DOCUMENT RESUME

ED 293 843 TM 011 198

AUTHOR Gabriel, Roy M.; Estes, Gary D.
TITLE An Evaluation of the Effectiveness of the Prescriptive Reading Inventory Reading System in Seattle Public Schools, 1983-84.
INSTITUTION Northwest Regional Educational Lab., Portland, OR. Assessment and Evaluation Program.
PUB DATE Nov 85
NOTE 146p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS Competency Based Education; *Curriculum Evaluation; *Educational Assessment; Elementary Education; *Language Arts; Mastery Tests; Program Evaluation; *Reading Achievement; Reading Instruction; Reading Materials; Reading Programs; Reading Tests
IDENTIFIERS California Achievement Tests; *Prescriptive Reading Inventory; *Seattle Public Schools WA

ABSTRACT

The adoption by Seattle (Washington) Public Schools of the Prescriptive Reading Inventory Reading System (PRI-RS) was a decision driven by the district's commitment to effective schooling. The PRI-RS is a criterion referenced approach to the instruction and assessment of reading skills, and it consists of five levels, A through E, coordinated with grade levels. The evaluation reported here consists of three major activities: (1) a statistical analysis of a district-wide data base, including demographic PRI use, and achievement data on more than 31,000 students; (2) structured interviews with principals and teachers in six schools in the district; and (3) an analysis of curricular match among PRI, California Achievement Test (CAT), and school reading curricula. Results indicate that: (1) the use of the PRI was highest for the early grades and significantly lower in Grades 7 and 8; (2) the average student in Grades 1-6 took five to six mastery tests before Spring CAT testing; (3) the frequency of PRI use was not a significant determinant of reading or language arts achievement gains; (4) quality PRI use was associated with achievement gains in reading or language arts; (5) PRI instructional materials contain many good items, but require extra work to implement; (6) effective classroom management strategies are critical in using the system well; and (7) the TRACER reporting system is burdensome and error-filled. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

THE Northwest Regional Educational Laboratory

PROGRAM REPORT

ED293843

An Evaluation of the Effectiveness of the Prescriptive Reading Inventory Reading System in Seattle Public Schools 1983-84

Dr. Roy M. Gabriel
Dr. Gary D. Estes

Evaluation & Assessment Program
Northwest Regional Educational Laboratory

November 1985

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Roy M. Gabriel

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Northwest Regional Educational Laboratory
300 S.W. Sixth Avenue
Portland, Oregon 97204

TM 011 198

BEST COPY AVAILABLE

An Evaluation of the Effectiveness of the Prescriptive Reading Inventory Reading System in Seattle Public Schools 1983-84

Dr. Roy M. Gabriel
Dr. Gary D. Estes

Evaluation & Assessment Program
Northwest Regional Educational Laboratory

November 1985

Acknowledgments

As referenced throughout this report, the authors had the cooperation and participation of an almost indeterminable number of Seattle Public Schools district and building level staff during the design and conduct of this evaluation study. Some were particularly noteworthy, and deserve at least the following recognition here.

Dr. Wayne Foley, of the Curriculum and Instruction Services department, served as our liaison throughout the study. His guidance as to the critical evaluation issues, reactions to our proposals, and direction to the appropriate district resources in several phases of the study were invaluable.

Dr. Nicholas Stayrook, of the Evaluation Services department, provided insightful and constructive reactions to the initial design of the study, and supervised the construction of the student data base in a timely and accurate manner.

Dr. Bernadette Cole-Slaughter, of the TRACER Reporting office, provided detailed information about that reporting system, a significant component of the system being evaluated in this study.

Dr. Alice Houston, Assistant Superintendent of Curriculum and Instruction, gave us important directions and ongoing perspective as to the importance of the evaluation study in the context of current issues in the district.

Several other district staff participated in meetings and briefings during the design stage of the study. Principals and teachers from six schools in the district took time out from their busy schedules to assist us in the interview activity in the study. School board representatives gave us helpful reactions in the early stages of the study. To all of these, we express our appreciation. It is our hope that the evaluation reported here is a useful one in the important decision-making processes ahead.

Contents

		Page
	Executive Summary	i
I.	Introduction	1
II.	Study Design	6
III.	The Prescriptive Reading Inventory Reading System (PRI/RS)	10
IV.	Question 1: Has the Use of the PRI/RS Improved Reading Achievement of Students in Seattle Public Schools?	17
V.	Question 2: What are some Effective Practices in PRI Implementation at the Building Level?	85
VI.	Question 3: Does the PRI/RS Curriculum Match that of the SPS and CAT?	97
VII.	Summary	105
	Appendixes:	110
	A. Study Design	
	B. Interview Guide and Schedule	

Executive Summary

The evaluation reported here consisted of three major activities--a statistical analysis of a districtwide data base including demographic, PRI use and achievement data on over 31,000 students; structured interviews with principals and teachers in six selected schools in the district; and an analysis of the curricular match among the PRI, CAT and the SPS Reading curriculum.

A brief summary of findings detailed in the Evaluation Report are as follows:

Levels of PRI Use

Use of the PRI was highest at the early grades, and significantly lower in grades 7 and 8.

In grades 1-6, the average student took five to six mastery tests before Spring CAT testing.

Impact on Achievement

Frequency of PRI use was not a significant determinant of Reading or Language Arts achievement gains.

Quality PRI use was associated with significant achievement gains in Reading or Language Arts.

Both of the above findings were consistent across all ethnic groups and socioeconomic levels.

Teacher Perceptions

PRI instructional materials contain many good ideas, but require extra work to implement.

Effective classroom management strategies are critical in using the system well.

The TRACER reporting system is burdensome and error-filled.

Chapter I.

Introduction

The adoption of CTB McGraw Hill's Prescriptive Reading Inventory Reading System (PRI/RS) in Seattle Public Schools (SPS) was a decision driven by the district's commitment to Effective Schooling, and one involving an array of vested interest groups. The authors of this evaluation believe the history of this adoption process is a relevant consideration in interpreting the findings of the current study. This history is detailed in other reports commissioned by the district (e.g., the PRI/RS Review Panel report submitted in Fall, 1984), so only a brief synopsis of the critical events and issues will be included here.

The need for an instructionally sensitive, ongoing assessment system was articulated by the Seattle School Board in response to the district's Effective Schools Seminar report in February, 1982. Such a system was seen as responsive to at least 2 of the 12 effective school characteristics outlined in that report: frequent monitoring of student progress and early identification of learning difficulties. In addition, board policy on testing stipulated the need for periodic diagnostic testing to assess student attainment of local curriculum objectives, as well as the use of more broadly based norm referenced tests to provide national comparisons for the achievement of SPS students.

The district considered developing its own criterion referenced testing system to meet these needs, but ultimately decided that the adoption of a commercially produced test was more beneficial. Following review of a number of existing criterion referenced assessment systems, the PRI/RS was selected.

Pivotal considerations included its ready availability, its record of use in other districts across the county, and its congruence with the California Achievement Test (CAT), the norm referenced test used by the district.

Once the adoption decision was made, a series of orientation and training sessions was held within the district. As noted in the 1984 district PRI/RS Review Panel report, the sequence was as follows:

Feb., 1983 -- An introduction to the system was provided for all members of the test committee. The full day workshop was attended by building principals, teachers, representatives from the school board, and central office staff.

April, 1983 -- A series of workshops for all elementary staff and selected secondary staff was held. School board and community members were also invited.

August, 1983-- Training by PRI/RS consultants was given to district central office staff.

-- Orientation and training was provided by PRI/RS consultants for elementary and middle school teachers and administrators. Community members were also invited.

Oct., 1983 -- Assistance in the interpretation and use of data from the diagnostic test portion of the system was provided to all elementary and middle schools by PRI/RS consultants. A make-up inservice for those building staff who missed the August training session was held.

Nov., 1983 -- A training session on the TRACER reporting system was conducted for teachers and building level administrators by the district TRACER system coordinator.

In addition to these scheduled training sessions, district staff provided ongoing assistance in the implementation of the system during the school year.

The need for a thorough evaluation of this new instructional/assessment system was seen by the district early in the adoption process. As formulated in a memo from the district superintendent, the evaluation of the PRI/RS would consist of two separate efforts. First, a process evaluation was to be conducted. Its purpose was to determine how well the PRI/RS was being implemented throughout the district. Secondly, a product evaluation was to determine the effects of the PRI/RS on the reading achievement of SPS students.

A PRI/RS advisory panel, whose membership included district and nondistrict staff, was formed in May, 1984. This review panel was commissioned by the district superintendent to conduct the process evaluation of the initial year's implementation of the PRI/RS. Its report consisted primarily of the results of a mailed survey to a random sample of elementary and middle school teachers throughout the district, and all elementary and middle school principals. The panel forwarded 12 recommendations based on its findings. Common themes among the 12 were the need for more training in the use of the PRI/RS at all levels, closer monitoring of its implementation, and several concerns for the computerized TRACER reporting system. Paraphrased, these recommendations were as follows:

1. The School Board should reaffirm its mandate to implement the PRI/RS district wide and its intent to monitor its use and evaluate its impact.
2. Training in the PRI/RS should be provided to principals and supervisors to support teachers' use and parents' understanding of the system.

3. Elementary and middle school supervisors should closely monitor the implementation of the PRI/RS in their assigned buildings.
4. A PRI/RS User's Group should be established to share ideas on effective use of the PRI/RS.
5. Special attention should be focused on the use of the PRI/RS Locator Test and the IOI diagnostic test at kindergarten and first grade.
6. The TRACER manual should be revised and expanded.
7. The amount of information teachers supply on TRACER data collection forms should be reduced to ease the reporting burden and minimize error.
8. Inservice training for teachers and principals on the TRACER system is needed.
9. Two types of reports considered useful by teachers should be added to the TRACER system.
10. In reporting PRI/RS results to parents, teachers should report how well a student has performed and suggest techniques for helping their child further.
11. A study should be done to determine why the Materials Resource File is not being used as expected.
12. Special attention should be paid to PRI/RS uses with students having special needs, e.g., low ability teenager readers.

District staff submitted their responses to each of these, concurring with virtually all of the Panel's recommendations. Their responses detailed the procedures they would follow (many of which were already in progress) to implement recommended changes/additions to current PRI/RS use in the district.

The findings of this study were seen as informative to the concerns for some of the implementation issues surrounding the use of the new system, as perceived by building level staff. The second evaluation study, to investigate the impact of the PRI/RS on student achievement, was commissioned by the district in Spring, 1984. Following a review of a proposal for such a study which was deemed unacceptable by the district, the Northwest Regional Educational Laboratory (NWREL) was contacted for assistance.

Preliminary discussions between NWREL evaluation staff and district curriculum and evaluation staff were held at SPS in September, 1984. NWREL staff, the authors of this report, then prepared a draft study plan with cost estimates for district consideration. The plan was reviewed favorably, minor revisions were made, costs were adjusted and a contract between SPS and NWREL was signed in October, 1984. One or both of the authors visited the district on four more occasions through January, 1985, culminating in the approval of the proposed study design for the conduct of the evaluation. These meetings included contacts with the district curriculum staff, research and evaluation personnel, a subcommittee of the school board, data processing staff, and community members. The approved study design was presented by the principal author of this report at a regularly scheduled meeting of the Seattle School Board, March 13, 1985. A copy of the study design is included as Appendix A of this report.

Chapter II.

Study Design

The process of designing this evaluation study included the following activities undertaken by the authors:

Interviews with Seattle district staff.

A study of the SPS Reading and Language Arts curricula.

Telephone interviews with CTB/McGraw Hill staff.

A study of PRI/RS materials.

Telephone interviews with district and school staff from other districts across the country using the PRI/RS.

A review of previous studies of PRI use. These included studies conducted within SPS, studies done by other districts using PRI, and studies done by CTB/McGraw Hill.

Interviews with SPS staff were conducted from the initial discussions in September, 1984 through the approval of the design in January, 1985. Importantly, a variety of audiences within the district was included in the design phase. Staff members from the areas of curriculum, research and evaluation, and data processing were consulted. A subcommittee of the School Board and a parent representative were also involved in these activities.

PRI/RS materials obtained from CTB McGraw Hill and SPS were studied and are summarized in the following chapter of the current report. Telephone interviews with CTB/McGraw Hill staff produced many useful technical reports regarding the PRI/RS, as well as nominations of districts around the country who had been using the system. These districts were contacted by the authors, and relevant experiences in the use of the PRI/RS were discussed in detail. Districts contacted by the authors included Atlanta, Georgia; Akron, Ohio;

Bakersfield, California; Memphis, Tennessee; Oklahoma City, Oklahoma; and Wilkesborough, North Carolina. These agencies represented as much as eight years experience with the PRI/RS, and provided valuable insights to the authors in the formulation of the study design. Oklahoma City was the only one which had conducted formal studies of the effectiveness of the PRI/RS, however. These studies, already well known to SPS staff, were obtained and studied by the authors. Their results provide some reference for interpreting the results of the current study.

The final study design included three major study activities in addressing the major questions of the district regarding the effectiveness of the PRI/RS. These activities were as follows:

Analysis of Student Data Base -- The major evaluation activity in the current study involved indepth statistical analyses of a district-wide data base containing individual student records of demographic information, PRI/RS use and performance, and CAT scores across two years. The specifications for the data base were developed by the authors in the design of the study to include data elements pertinent to all study questions. A descriptive summary of the 100 variables on each of over 31,000 student records in this data base is given in Table II-1. A more detailed specification of a student record in the data base is given in the Study Design in Appendix A.

Teacher Interviews -- In addition to the "bottom line" summative evaluation questions of the study, there were important implementation questions as well. The PRI Review Panel report addressed many of these in the mailed survey they conducted in the spring of 1984. The current evaluation sought to expand on these findings and pursue others through personal interviews with teachers and principals in selected schools in the district.

Curriculum Assessment -- The match between the PRI/RS curriculum with that of Seattle Public Schools is a critical consideration in the evaluation of the effectiveness of that system. If the instructional activities and assessment tools of the PRI/RS address different reading skills than those of the SPS curriculum, it casts doubt on prospects for that system to be effective in the district. A third curricular component important in this evaluation is the content map of the district's norm referenced test, the California Achievement Test. If the reading objectives it tests at each grade level do not match those of the PRI/RS and the SPS curriculum, it would be an insensitive instrument to use in determining whether the PRI/RS was effective in raising students' reading achievement.

These study activities were designed to address the major study questions in the evaluation of the effectiveness of the PRI/RS. As formulated by the authors, based on extensive interviews and discussions with SPS district staff, they are:

Question 1: Has the use of the PRI/RS improved reading achievement of students in Seattle Public Schools?

Question 2: What are the important implementation issues related to PRI use at the building level? What are some effective practices in its implementation at the building level?

Question 3: Does the curriculum map of the PRI/RS match that of the SPS and the CAT?

These three questions are addressed in subsequent chapters of this report.

Table II-1

A Summary of a Student Record in the
Data Base

Demographic Data

Student, School and Teacher ID
Grade Level in 1983-84
Gender
Free Lunch Status in 1983-84 (SES measure)
Bilingual Fluency in 1983-84
Chapter 1 Participation in 1983-84
Special Education Participation in 1983-84
Ethnic Origin

PRI Performance

For each of the 5 PRI levels, and totalled separately for the period before and after Spring CAT testing:

Date IOI diagnostic test administered
Number of PRI objectives mastered on the IOI
Number of mastery tests administered
Number of objectives mastered via mastery tests
Number of objectives mastered via teacher rating

District-wide Testing (CAT) data

For Spring, 1983 (pre-PRI) and Spring, 1984, raw scores, scale scores, and NCEs for the following subject areas:

Reading Vocabulary
Reading Comprehension
Total Reading
Language Arts Mechanics
Language Arts Expression
Total Language Arts
Math Computation
Math Concepts and Applications
Total Math

Form and level of CAT administered.

Chapter III.

The Prescriptive Reading Inventory/Reading System

The Prescriptive Reading Inventory (PRI), published by CTB/McGraw-Hill, is a criterion referenced approach to the instruction and assessment of reading skills. The system spans the kindergarten through high school grade levels, and is available in graded or ungraded formats. Since Seattle Public Schools adopted the graded alternative (System 1 in CTB terminology), this chapter will concern itself only with it.

The PRI/RS consists of five levels, A-E, coordinated with grade levels as shown in Table III-1. Each level includes a different number of reading objectives, also shown in Table III-1. The PRI materials are designed and organized by these objectives. The overall package of materials, as adopted by SPS, consists of the following components at each level:

Locator Test -- a brief test of vocabulary and comprehension used to determine the appropriate level of PRI materials and tests for each student.

Instructional Objectives Inventory (IOI) -- a diagnostic test covering all objectives at a given PRI level. Each objective is assessed with four items, and the resultant score is designed to be used to classify the objectives as having been mastered, needing review, or not mastered. This information is to assist teachers in planning instructional programs for their students.

Instruction Materials Kit -- a compendium of resources teachers may use in designing instructional activities for their students. They are organized by PRI objective, and correlated with a variety of basal curriculum materials.

Mastery Test -- a short, six-item test of each PRI objective, designed to be administered when the student has completed the instructional activities related to that objective and/or appears ready to demonstrate mastery.

TRACER -- a computerized instructional management aide which provides a wide variety of reports back to teachers based on the mastery test results they submit. It is designed to be a quick turnaround system, giving teachers up-to-date information on what objectives each of their students are currently working on, what they have completed, and what work is planned.

Table III-1

PRI/RS Levels, Recommended Grade Levels, and Number of Instructional Objectives

PRI/RS Level	Grade Level	No. Instructional Objectives
A	K,1	13
B	1,2	39
C	2,3	43
D	4,5,6	42
E	7,8	34

Full implementation of the PRI/RS involves use of all of the components described above. At the beginning of the school year, students are administered the Locator test to place them in the appropriate PRI/RS level. Although each PRI level is associated with a recommended grade level or range of grade levels, the system allows for variations for very high or low achieving students. Following the Locator test, the designated level of the IOI is administered. These results provide a map of student mastery/nonmastery on all objectives at that PRI level. Once teachers know which objectives their students have mastered, and which ones need instruction, they can begin planning the instructional program for their students. The PRI/RS instructional materials include teacher resource guides, student exercises and tutor activities to supplement teachers' instructional programs. When students have completed activities for a given objective, or when the teacher feels the students are ready, a six-item mastery test for that objective is administered. The results of this test are sent to TRACER, and the student's PRI/RS record is updated. This cycle of instruction and testing by objective continues throughout the school year. If all objectives in the given PRI

level are mastered, students may move to the next level. Their cumulative records can follow them to the beginning of the next school year, and instructional programs may be designed based on their history of progress through the PRI system.

The content of the PRI materials varies by level, but is organized in a four-tiered hierarchy as shown in Figure III-1. Across all levels, there are 171 Instructional Objectives included in the system. Many are taught and tested at more than one grade level. These objectives are classified into 75 slightly more general Category Objectives, which are further subsumed within nine Skill Areas. Finally, these skill areas are themselves organized within four Skill Clusters. A sample content specification at each of these four tiers is shown in Figure III-1. A tally of the number of objectives within each skill area included in each PRI level is shown in Table III-2. The flow of content across PRI and grade levels is clearly displayed in this way. Oral Language skills are emphasized at the early grades; Word Attack & Usage and Comprehension skills pick up through the middle grades; and Applications are emphasized in the junior high school years. The overlap in objectives taught across levels increases once the oral skills are completed.

The test items contained in both the IOI and mastery tests are primarily four-option multiple choice. Two notable exceptions include orally administered questions for pre-K and kindergarten students using cassette tapes, and the use of reading passages to assess the comprehension skills of older students. CTB test developers indicate that, at all levels, items are at a readability and difficulty level commensurate with the instructional objective at that grade level; and represent population characteristics (gender, age, socioeconomic status, ethnic group, urbanicity) in the proportions they occur in society.

Figure III-1

Content Hierarchy in PRI/RS

	Total Number in PRI/RS	Example
SKILL CLUSTER	4	Wood Attack and Usage
SKILL AREA	9	Vocabulary
CATEGORY OBJECTIVE	75	Word Matching
INSTRUCTIONAL OBJECTIVE	171	Antonyms

Table III-2

Number of Instructional Objectives in each Skill Cluster and Area by PRI/RS Level

Skill Cluster	Skill Area	Level A Grades K,1	Level B Grades 1,2	Level C Grades 2,3	Level D Grades 4-6	Level E Grades 7-8
Oral Language	Oral Language	5	7			
	Oral Comprehension	4	5			
Word Attack & Usage	Word Analysis	4	7	8	5	
	Vocabulary		5	6	4	4
	Word Usage		8	9	8	7
Comprehension	Literal Comprehension		3	5	4	3
	Interpretive and Critical Comprehension		4	5	7	8
Applications	Study Skills			1	5	3
	Content Area Reading	—	—	9	9	9
	Total	13	39	43	42	34

As described above, student mastery of the PRI/RS objectives is the learning contingency for progress through the materials. From a theoretical point of view, the criterion for mastery is an important issue which has shadowed the criterion referenced testing movement since its inception. The IOI covers each objective with four test items; mastery tests use six. Results for both are expressed in one of three ways: mastery, review needed, and nonmastery. The number correct needed for each is shown below for both tests:

		Mastery	Review	Nonmastery
IOI	(4 items/obj)	3 or 4	2	0 or 1
Mastery	(6 items/obj)	5 or 6	3 or 4	0, 1 or 2

These criteria, as specified by the PRI/RS authors and publisher and recorded by the TRACER, are used in the current evaluation study to represent mastery/nonmastery scores in the PRI/RS for all SPS students.

Table 4

"Critical Events" in PRI Study

Task	Due Date	NWREL Deliverable
Activity 4: Personal Interviews		
NWREL obtains teacher rosters from SPS	February 1, 1985	
NWREL identifies high, medium and low PRI users via TRACER analysis	February 15, 1985	
NWREL selects teacher and principal sample	February 22, 1985	
Interview schedule finalized	March 1, 1985	
Interviews conducted	March 22, 1985	
Progress Report of interview results to SPS	June 15, 1985	***
Final Report, synthesizing findings of all Progress Reports in all study activities	November 15, 1985	***

Chapter IV.

Evaluation Question 1: Has the use of the PRI/RS improved reading achievement of students in Seattle Public Schools?

In the design stage of this study, this question was by far, the single most important question to the district. Indeed, it is the classic summative evaluation issue -- the "bottom line" through which the effectiveness of the PRI/RS will be judged.

The authors address this question through the database at hand with all the familiar methodological caveats of field-based evaluation research. The data available are extensive and the sample size is very large, allowing the authors to use sophisticated analytical techniques with the precision to make strong statements about the statistical significance of the relationships detected. These factors get us no closer to the causal inference implied in the question above, however. Students in Seattle Public Schools may be reading better after using the PRI for a year for reasons unrelated or only tangentially related to the PRI/RS. For example, if the adoption of the PRI coincided with the adoption of a new basal reading series in all schools, the improvement may be due to the latter rather than the PRI. Or perhaps a drastic decline in enrollment in the district resulted in more favorable student/teacher ratios in the classrooms, allowing more time for individual instruction. Or perhaps the widespread adoption of a system as comprehensive as the PRI simply focused more attention on reading throughout the district, and the difference is due to this increased attention (and, likely, resources) rather than the substance of the PRI/RS itself. Or...

These kind of provisional caveats are seldom satisfying, and often irritating, to policy makers. The authors have designed their analyses to minimize the likelihood of alternative explanations to observed differences in

reading achievement. Breaking down PRI Use into three levels--high, medium and low is an example. If high PRI users read better than medium or low PRI users, this is highly suggestive of a close association between PRI Use and student achievement, even though scientific rigor would not deem it as proof of a causal connection. What if the better readers naturally fall into the high use group anyway? Is it not a foregone conclusion that they will show higher reading scores on the CAT? This potential bias is dealt with statistically. One of the analyses conducted by the authors statistically adjusts differences on the 1984 CAT scores among the three PRI Use groups for any pre-existing differences on their 1983 (pre-PRI) CAT scores. Finally, what if the increase in achievement is due to the additional attention given the students, rather than the nature of the PRI materials themselves. The authors contend that this increased attention is part and parcel of the PRI/RS intervention, and does not diminish the significance of its impact.

In designing our approach to this crucial question, two distinct analytical approaches were devised--Multiple Linear Regression, and the Analysis of Covariance. They differ in their statistical operations, but address the same general hypotheses. Multiple Linear Regression (MLR) takes a correlational approach to assessing the relationship of PRI use to reading achievement. The Analysis of Covariance (ANCOVA) employs comparisons of average reading scores among relevant subgroups of students, after statistically equating those groups for pre-PRI achievement differences. In this application, the two methods differ primarily in the detail with which they represent the variables of interest. For example, our MLR represents PRI use for each student as the specific number of times that student took a mastery test; our ANCOVA, on the other hand, simply classifies each student as

a high, medium, or low PRI user depending upon the number of mastery tests he or she took. While MLR offers the more detailed statistical assessment of the PRI Use/Reading Achievement relationship, ANCOVA offers a more easily interpreted approach to the group comparisons of interest in this study (e.g., PRI Use groups, ethnic groups, SES groups).

The authors view MLR and ANCOVA as complementary analytical approaches to the principal question in this evaluation. They will be presented in separate sections of this chapter, and their results will be synthesized in a summary following their presentation. The MLR analysis will be presented first, because in addition to addressing the primary evaluation question, it offers results which relate to the potential usefulness of the ANCOVA approach.

The Multiple Linear Regression Approach

The Spring 1984 CAT Reading scores are the principal criteria of interest in this evaluation. These are taken from the district-wide testing program conducted near the end of the first year of PRI/RS implementation. The purpose of the regression analysis reported here is to determine to what extent these scores are a function of varying degrees of PRI use and mastery, or are simply a reflection of students' previous reading achievement status. Specifically, 1984 CAT reading scores were regressed against the following four predictor variables:

IOI Score -- students' scores in the Fall 1983 administration of the Instructional Objectives Inventory.

PRI Use -- the number of mastery tests administered during the 1983-84 school year prior to Spring CAT testing.

PRI Mastery -- the number of PRI objectives mastered via mastery testing during the 1983-84 school year prior to Spring CAT testing.

1983 CAT Reading Score -- the reading achievement score from district-wide testing the previous school year.

The first three predictor variables are reflections of PRI Use during its first year of implementation. The IOI score functions as a diagnostic "pretest" at the beginning of the school year. The PRI Use variable is a tally of the number of times a student took a PRI Mastery test. It is used here as an indicator of how heavily the PRI/RS was used with each student, and is a crucial consideration in addressing the current evaluation question. The PRI Mastery variable is a tally of the number of objectives a student mastered in the mastery testing. It is an indicator of successful use of the PRI/RS. Finally, the 1983 CAT Reading scores were included to represent students'

previous reading achievement status. In the multiple regression analysis, inclusion of this as a predictor variable effectively partials out the influence of previous reading achievement, and assesses the relationship of the PRI variables to Spring 1984 Reading scores over and above this important factor.

The analysis described above was also conducted for Language Arts and Mathematic achievement scores. Language Arts was included because some of the PRI instruction relates to the Language Arts curriculum, particularly in the higher grade levels. Math achievement was included to assess a possible general educational effect of the PRI/RS--one which would override specific parallels with instructional content.

Correlational analyses using MLR were conducted separately for each grade level so as not to obscure any grade to grade differences. Results of these analyses are presented in Table IV-1. Simple correlations between each of the four predictor variables and 1984 CAT Reading (or Language Arts or Math) scores are shown, as are the multiple correlations of the four variables combined with 1984 CAT scores. At all grades and subject matters, the predictor variable with the highest correlation with CAT 1984 scores is the CAT 1983 scores. These correlations range from .64 to .88 in Reading, from .62 to .84 in Language Arts, and from .66 to .88 in Math. The strength and consistency of these relationships are not surprising. They compare very favorably to the test publisher's test-retest reliability statistics, particularly considering the longer time interval between these testing points (1 year) and those typically used in test norming studies (2-4 weeks).

Previous achievement level is virtually always the strongest predictor of current achievement level. The question of interest here is, once we have taken into account a student's prior achievement level, to what extent is

Table IV-1

Simple (and Multiple) Correlations of all Predictor Variables
With 1984 CAT Reading, Language Arts and Math Scores by Grade Level

Grade:	2	3	4	5	6	7	8
1984 CAT Reading							
1983 CAT Reading	.64	.75	.73	.85	.85	.85	.88
IOI	.56	.36	.25	.32	.25	.22	.32
PRI Use	.12	.07	.05	-.01	.09	.00	-.09
PRI Mastery	.19	.02	-.01	.01	-.06	.04	-.04
(Multiple R)	(.72)	(.76)	(.74)	(.85)	(.85)	(.85)	(.88)
1984 CAT Language Arts							
1983 CAT Lang. Arts	.62	.66	.66	.78	.80	.81	.84
IOI	.48	.33	.25	.35	.30	.24	.33
PRI Use	.17	.02	.05	.02	-.04	.04	-.08
PRI Mastery	.23	.01	.02	.05	.00	.08	-.03
(Multiple R)	(.67)	(.67)	(.67)	(.79)	(.80)	(.81)	(.84)
1984 CAT Math							
1983 CAT Math	.68	.67	.69	.80	.82	.84	.88
IOI	.52	.30	.22	.33	.26	.24	.36
PRI Use	.11	.03	.07	.02	-.08	.03	-.08
PRI Mastery	.14	-.02	.04	.05	-.04	.08	-.03
(Multiple R)	(.72)	(.68)	(.70)	(.80)	(.83)	(.84)	(.88)

his/her current achievement a function of experience with the PRI system? By contrasting the simple correlation of CAT 1984 scores and CAT 1983 scores with the multiple correlation of CAT 1984 scores and all predictors (i.e., CAT 1983 scores and PRI variables), we determine the additional predictability of current reading achievement due to PRI experience. At all but grade 2, the difference between these two correlations never exceeds .01--indicating almost no additional predictability in current achievement due to PRI experience.

A more statistical treatment of this contrast is shown in Table IV-2. The previously discussed correlations are squared so as to represent the more (statistically) conventional notion of shared variance between criterion and predictor variables. The difference between these two (designated "Residual" in Table IV-2) represents the additional proportion of variance in the criterion variable accounted for by the PRI variables. The F statistic given in the table tests whether this Residual is significantly greater than zero, i.e., is the additional predictability of reading achievement due to PRI experience statistically significant?

The results of the statistical comparisons parallel those of the visual analyses. At all but grade 2, the Residual values are very small. The F statistics all attain conventional levels of statistical significance, but this is largely due to the extremely large sample sizes in this analysis. The authors maintain that increases in shared variance due to PRI experience of one or two percent, compared to the 50-70 percent variance accounted for by prior achievement status, are relatively minor. In methodological parlance, these Residuals are considered statistically significant, but, in our view, not educationally significant.

Table IV-2

Squared Correlations and Residuals in
Multiple Regression Analyses

Grade:	2	3	4	5	6	7	8
Reading							
1983 CAT Reading Only	.41	.56	.53	.72	.72	.71	.78
All Predictors	.52	.58	.54	.73	.73	.72	.78
Residual	.11	.02	.01	.01	.01	.01	.00
F ratio	105.77	22.06	19.23	19.77	8.33	6.67	4.17
Language Arts							
1983 CAT Language Arts Only	.38	.43	.66	.60	.64	.65	.71
All Predictors	.45	.45	.67	.62	.64	.65	.71
Residual	.07	.02	.01	.02	.00	.00	.00
F ratio	90.91	21.21	28.57	29.63	9.52	3.95	9.38
Math							
1983 CAT Math Only	.46	.44	.47	.63	.68	.70	.77
All Predictors	.52	.46	.49	.64	.68	.70	.77
Residual	.06	.02	.02	.01	.00	.00	.00
F ratio	71.43	21.88	14.52	16.67	11.76	2.19	6.67

These regression analyses were conducted within each grade/subject combination, but across relevant subgroups in the Seattle Public Schools population (e.g., ethnic groups, SES). The correlational relationships reported are, therefore, aggregate in nature and insensitive to possible differences among these subgroups. These factors are better represented in the second analytical approach to the principal evaluation question of the study--the analysis of covariance (ANCOVA). The high correlations reported between 1983 and 1984 CAT scores play an important role in these analyses. Using the prior achievement status as a covariate not only equates the subgroups for pre-PRI differences in achievement status, but also reduces the error variance in subsequent statistical tests of subgroup differences. The size of the correlations between the 1983 and 1984 CAT scores indicates that this reduction will range from 50-70 percent, allowing for very precise comparisons of PRI Use levels, ethnic groups, and SES levels.

The Analysis of Covariance Approach

An alternative method for assessing the effect of PRI/RS use on student achievement is to compare students of varying levels of use of the PRI/RS on their CAT Reading scores. If the PRI/RS has had an effect on students' reading achievement, "high", "medium" and "low" PRI users should differ in their reading scores after using the system. Since these groups of students may also be at different reading achievement levels independent of PRI use (e.g., "high" users may be better readers who progress through the system more rapidly), their previous reading achievement scores are used to adjust the current scores for prior differences. Once these pre-existing differences are statistically accounted for, resultant comparisons of adjusted means in the ANCOVA are valid indicators of the effects of varying PRI Use on subsequent reading achievement.

There are three important groupings of students for this analysis--levels of PRI Use, Ethnic Group, and Socioeconomic status (SES). PRI Use is represented by three levels--high, medium, and low. Determination of these levels was done on the basis of actual frequencies of use, and is reported later in this section. The ethnic group is represented in five categories--Asian, Black, Hispanic, Native American, and White. Socioeconomic status consists of two categories of students--those who qualify for the free or reduced lunch program (low SES) and those who do not (high SES). These three groupings constitute a three factor ANCOVA design. The design is displayed graphically in Figure IV-1.

Figure IV-1

Three Factor Design for MANCOVA

Ethnic Group

	Asian		Black		Hispanic		Native American		White	
PRI Use	Low SES	High SES	Low SES	High SES	Low SES	High SES	Low SES	High SES	Low SES	High SES
High										
Medium										
Low										

The complete analysis of this three factor design includes tests of seven effects--termed main effects and interaction effects--related to these factors. Of these, four are of primary interest in this evaluation:

PRI Use Main Effect -- Compares adjusted mean CAT reading scores among high, medium and low PRI Use groups. This test is done across all other factors in the design.

PRI Use by Ethnic Group Interaction -- compares the pattern of differences detected in the PRI Use main effect across the five ethnic groups. This is to determine whether the general pattern of differences in CAT scores among PRI Use groups is the same or different for each ethnic group.

PRI Use by SES Interaction -- is analagous to the interaction described above. It determines whether the PRI main effect is the same for high and low SES students.

PRI Use by Ethnic Group by SES Interaction -- compares the pattern of reading score differences detected in the PRI Use main effect across combinations of the Ethnic and SES factors. It will determine, for example, whether the PRI Use main effect is the same for high SES White and low SES White students.

Summarily, the PRI Use main effect addresses the primary evaluation question of the study. The interactions assess the generalizability of this effect (or non-effect) across ethnic groups and SES levels.

As in the regression analysis, the ANCOVAs were conducted separately for each grade level so as not to obscure any effects present in some grades but not others. In addition, the three subject areas--Reading, Language Arts and Math--are analyzed simultaneously in a multivariate analysis of covariance (MANCOVA) to minimize redundancy in assessing the effects of PRI Use, and to maximize statistical power. Differences in CAT Reading, Language Arts and Math scores are also analyzed separately within this multivariate context to maintain interpretive clarity.

In order to proceed with the comparisons of CAT scores among the various student groups of interest, some critical distinctions must be made. How do we define high, medium, and low levels of use of the PRI/RS? Do these standards vary by grade level or student groups? The authors view these questions as important in themselves, and introduce a rather detailed and important empirical digression at this point.

Instructional Objectives Inventory

The use of the PRI/RS began with the administration of the Instructional Objectives Inventory (IOI) in Fall, 1983 to the entire elementary and middle school population in Seattle Public Schools. The IOI was administered initially in late October, and again in subsequent months as needed for new students or students requiring a different level than that given originally. Scores reported indicated the number of objectives mastered for the level administered.

The results of the IOI are important to the implementation of the PRI/RS program because they provide information to teachers and building administrators as to which reading objectives need instruction or remediation. They provide a starting point in the planning of the instructional program. For this reason, it is important that the appropriate level of the IOI is administered to each student. Obtaining a perfect score, or a zero score, on the IOI provides virtually no information to the teacher. Should a student score at or near either of these extremes, a different level should be administered, one that is more in line with the student's functional reading level.

In the fall of the 1983-84 school year, the IOI was administered to more than 26,000 students in grades 1-8 in SPS. Across the entire district, several PRI levels were administered at each grade. Table IV-3 contains the proportion of students at each grade level taking each level of the IOI. The bracketed [ ] cells indicate the recommended PRI level(s) at each grade. At all grades, the largest percentage of SPS students took the recommended level

Table IV-3

Number and percent* of students taking each level of the IOI at each grade

		Level A	Level B	Level C	Level D	Level E
Grade 1	N	3198	1762	71	21	31
(N=3620)	%	88	49	2	1	1
Grade 2	N	680	2234	955	162	43
(N=3102)	%	22	72	31	5	1
Grade 3	N	161	1134	1812	744	82
(N=3009)	%	5	38	60	25	3
Grade 4	N	89	555	1259	1753	334
(N=3094)	%	3	18	41	57	11
Grade 5	N	62	342	811	2087	726
(N=3136)	%	2	11	26	67	23
Grade 6	N	36	170	559	1816	1411
(N=3138)	%	1	5	18	58	45
Grade 7	N	26	209	439	1339	2089
(N=3507)	%	1	6	12	38	60
Grade 8	N	50	167	360	1007	2551
(N=3671)	%	1	4	10	27	69

* Percentages do not add to 100 within each grade because many students took more than one level of the IOI.

of the IOI, although there are significant proportions of students taking alternative levels. For example, at grade 3, 60 percent of the students took Level C (the recommended level), while 38 percent took Level B and 25 percent took level D. Many of these are the same students. If one were to add the percentages shown in Table IV-3 for each grade, they would add to over 100% -- evidence that many students took more than one IOI level.

Descriptive statistics on SPS students IOI performance are shown in Tables IV-4 and IV-5. These statistics are presented for each PRI/RS level on which at least 10% of the students at that grade level were tested (see Table IV-3). The mean and standard deviation of the number of objectives mastered on the IOI are shown in Table IV-4. Means are also expressed as percentages of the total number of objectives tested in each PRI/RS level. With the exception of grade 1, Level B, all grades show IOI averages in the 40-79 percent range for all levels at which at least 10% of the students were tested. At a group level, these results are in an ideal range from both assessment and program implementation perspectives. They are well above the floor and below the ceiling of the test, indicating reliable assessment. They also suggest a workable balance of mastered and unmastered objectives for prescribing instructional activities at each PRI level. We cannot, however, infer this to be the case for all students tested. The relatively high standard deviations suggest that individual student scores are widely dispersed from zero to the total possible for each PRI/RS level. This is illustrated by the grouped frequency distributions summarized in Table IV-5. To examine this further, students' IOI scores were grouped into four categories: 0-25 percent of the objectives mastered, 26-50 percent mastered, 51-75 percent mastered, and 76-100 percent mastered. These grouped

Table IV-4

Mean (M), Standard Deviation (S) and Percent (%)
IOI Objectives Mastered by Grade and PRI/RS Level[1]

		Level A (13 obj.)	Level B (39 obj.)	Level C (43 obj.)	Level D (42 obj.)	Level E (34 obj.)
Grade 1	M	8.11	7.13			
	S	3.15	9.83			
	%	62	18			
Grade 2	M	8.97	17.42	19.02		
	S	3.10	10.82	13.47		
	%	69	45	44		
Grade 3	M		21.52	19.87	17.76	
	S		10.59	13.32	11.76	
	%		55	46	42	
Grade 4	M		22.98	20.26	19.46	15.99
	S		10.78	12.15	11.59	9.89
	%		59	47	46	47
Grade 5	M		23.07	20.77	22.49	15.47
	S		10.76	13.09	11.77	10.91
	%		59	48	54	46
Grade 6	M			22.08	22.52	18.22
	S			13.50	11.42	10.88
	%			51	54	54
Grade 7	M			25.77	20.42	20.62
	S			10.79	10.86	9.63
	%			60	49	61
Grade 8	M			23.41	21.26	22.04
	S			11.56	10.71	9.04
	%			54	51	65

[1] Statistics are shown only for those PRI/RS levels on which at least 10% of the students at a grade level were tested.

Table IV-5

Grouped Frequency Distributions of each level of the IOI at each grade

	Percent Ojectives Mastered	Level A	Level B	Level C	Level D	Level E
Grade 1	0- 25%	10	70			
	26- 50%	25	17			
	51- 75%	40	8			
	76-100%	25	5			
Grade 2	0- 25%	8	27	32		
	26- 50%	14	33	20		
	51- 75%	30	26	31		
	76-100%	48	14	17		
Grade 3	0- 25%		16	29	32	
	26- 50%		25	22	33	
	51- 75%		37	32	34	
	76-100%		22	17	11	
Grade 4	0- 25%		16	23	27	25
	26- 50%		22	27	27	20
	51- 75%		32	35	31	39
	76-100%		30	14	15	16
Grade 5	0- 25%		13	26	19	30
	26- 50%		23	21	24	16
	51- 75%		32	35	32	38
	76-100%		32	18	25	16
Grade 6	0- 25%			24	18	23
	26- 50%			17	24	13
	51- 75%			35	35	38
	76-100%			24	23	26
Grade 7	0- 25%			11	21	15
	26- 50%			20	29	16
	51- 75%			42	36	35
	76-100%			27	14	34
Grade 8	0- 25%			18	19	12
	26- 50%			20	27	15
	51- 75%			41	40	33
	76-100%			21	14	40

distributions show the highest percentages of students in the middle two categories (26-75 percent mastery), thus suggesting a desirable range of difficulty within each IOI level administered to more than 10% of the students at a grade. Typically, there is not a large proportion of students in the lowest (floor) or highest (ceiling) categories. Again, an exception is at grade 1. Those students taking level A, the orally administered level, evidence useful results as previously described. Those students taking level B, however, are heavily clustered in the lowest category, i.e., less than 25 percent of the IOI objectives mastered. This "floor efect" on level B for first graders likely reflects the widest chasm between adjacent PRI/RS levels.

In summary, the IOI results reflect positively on both the suitability of the tests, and appropriate district implementation.

Mastery Tests

Once the results of the IOI are available, instructional programs are designed to remediate those objectives indicated to be in need of instruction. When these activities are completed, or when the teacher judges students to be ready, mastery tests are administered for that objective. The frequency of this mastery testing, as well as its results, is important in the current study. It is the frequency of the use of the mastery test materials which characterize students as high, medium, or low users of the PRI system. Further, it is important to distinguish between the frequency of use before and after the Spring districtwide testing, since it is these scores which serve as the criterion in evaluating the effectiveness of the PRI in the district.

As with the IOI, there was considerable evidence of students moving through more than one level of the PRI materials. It is less critical to report these results by PRI level, however, since the mastery testing is designed to be conducted continuously throughout the year. Table IV-6 presents descriptive statistics on the number of mastery tests submitted per student both before and after the Spring CAT testing (April 15) by grade level. Before Spring testing, grades 1-5 average five to six mastery tests submitted for each student, while grade 6 is approximately four. Middle school students, grades 7-8, show substantially less use - only two to three tests submitted per student. The pre-CAT period extends from early December, when IOI results became available to mid-April, the time when the CAT is administered districtwide. For the sake of interpretation, we will consider this to be a four month time period (deducting some time for Christmas holidays). The post CAT period is about two months. In other words, through grade six, tests are submitted at a rate at or better than one per month on the average. In grades 7-8 they occur only half as often. After Spring testing, the pace picks up considerably, particularly in the middle schools. Grades 7 and 8 students actually took more mastery tests in the two-month period after the CAT testing than in the four-month period before it. Earlier grades also evidence an increased rate of submissions after CAT testing, but none as dramatic as that in grades 7 and 8.

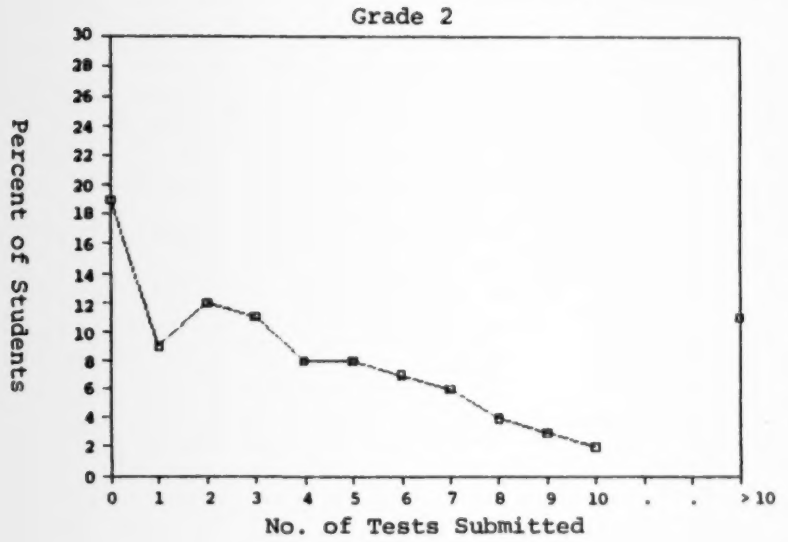
Descriptive statistics on the frequency of mastery testing also include standard deviations. In every case they are far larger than the means, suggesting a highly skewed distribution of mastery test use. Typically, over two-thirds of the students at a grade level exhibit below average use. A small proportion of students exhibit very high use, thus raising the overall

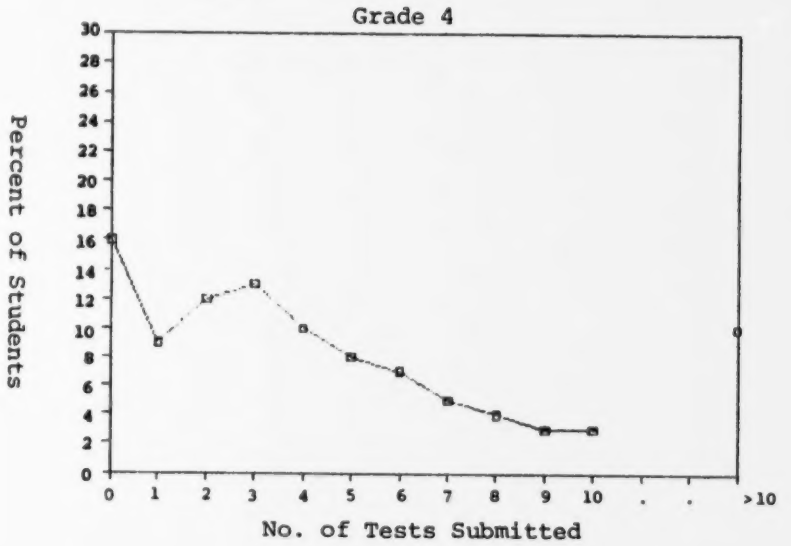
Table IV-6

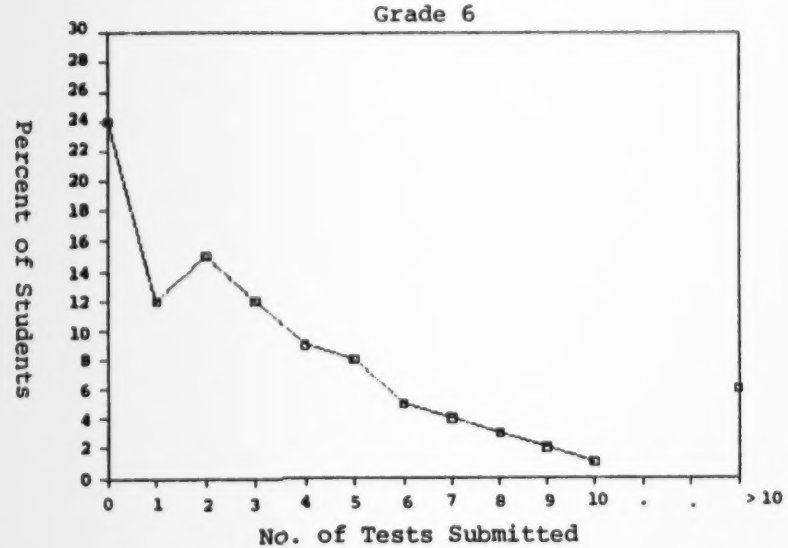
Mean (M), Standard Deviations (S) and Percent at 0 and above 10
Mastery Tests Submitted Before and After Spring Testing, 1984

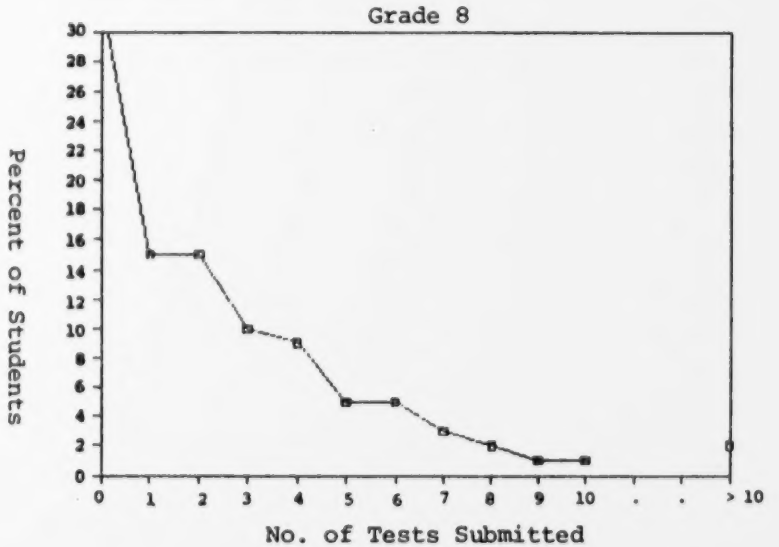
	Before Spring Testing				After Spring Testing			
	M	S	% at 0	% > 10	M	S	% at 0	% > 10
Grade 1	5.22	7.45	19%	12%	2.92	4.29	41%	6%
Grade 2	5.56	8.46	19%	13%	4.18	6.00	38%	12%
Grade 3	4.98	7.69	18%	8%	3.21	5.90	54%	9%
Grade 4	5.66	8.56	16%	13%	2.62	5.23	60%	8%
Grade 5	5.31	8.14	18%	11%	2.89	5.10	48%	7%
Grade 6	3.98	5.72	23%	7%	2.58	4.55	49%	5%
Grade 7	2.09	2.84	33%	2%	3.16	4.91	36%	7%
Grade 8	2.59	3.69	32%	3%	3.66	4.95	27%	8%

Figure IV- 2
Illustrative Frequency Distributions of Number
of Mastery Tests Submitted prior to CAT Testing









average. Figure IV-2 contains a plot of the number of mastery tests submitted at several grade levels. A substantial number of students (usually about 20 percent in grades 1-6, over 30 percent in grades 7-8) had no mastery tests submitted at all. A few students had an extremely high number of tests submitted. To give a fuller picture of mastery test use at each grade, the percent of students with no tests submitted, and the percent with more than 10 submitted before CAT testing are also presented in Table IV-6.

The number of objectives mastered by each student during the mastery testing is another important consideration in the use of the PRI. Descriptive statistics on this important outcome are presented in Table IV-7. They show a trend similar to that described above for the number of tests submitted. Their standard deviations exceed their means, yielding a highly skewed distribution of mastery. On the average, three to four objectives are mastered via mastery testing prior to Spring CAT administration through grade six. In grades seven and eight, the average is one to two. After Spring CAT testing, the rate of mastery increases along with the previously mentioned frequency of mastery. Again, this is most dramatic at grades 7 and 8 where students actually mastered more objectives after Spring CAT testing than before.

The division between PRI/RS mastery testing before and after Spring districtwide testing is most critical for the subsequent analysis of SPS students CAT reading achievement test scores as they relate to frequency of PRI use. High, medium and low users must be so classified before Spring testing if the latter is to accurately reflect effects of these varying levels of PRI/RS use. The data presented in Table IV-6 for the pre-CAT period is informative as to average levels of use of mastery tests at all grades. The

Table IV-7

Means (M), and Standard Deviations (S) of Number of Objectives Mastered before and after Spring Testing

	Before Spring Testing		After Spring Testing	
	M	S	M	S
Grade 1	3.28	4.79	1.71	2.85
Grade 2	3.59	5.41	2.83	4.16
Grade 3	3.04	4.53	2.23	4.33
Grade 4	3.15	5.32	1.75	3.67
Grade 5	3.00	5.23	1.72	3.13
Grade 6	2.47	4.59	1.48	2.81
Grade 7	1.13	1.97	1.70	2.50
Grade 8	1.34	2.32	1.94	2.57

classification of students into "high", "medium" and "low" user groups, however, requires more detailed statistical analysis. For this purpose, frequency distributions of mastery tests submitted prior to CAT testing were constructed. The authors tried to define these three levels such that: (a) they divided students in three roughly equivalent size groups within a grade, and (b) the same definitions could be used for all grades. The latter principle was of interpretive interest so that a high user at one grade would also be a high user at another grade. When comparing results across grades, this consistency seemed desirable.

Descriptive statistics in mastery test use already presented (Table IV-6) indicated the degree and pattern of PRI/RS use was markedly different in the middle schools (grades 7-8) from all earlier grades. Frequency distributions corroborated this interpretation, and the consistency principle mentioned above could not be achieved. Thus, we have two different classification rules for high, medium and low PRI/RS users - one for grades 1-6 (SPS elementary schools) and the other for grades 7-8 (SPS middle schools). These are as follows:

		Grades 1-6	Grades 7-8
PRI Use	High	5 or more mastery tests	3 or more mastery tests
	Medium	2-4 mastery tests	1-2 mastery tests
	Low	0-1 mastery test	0 mastery tests

Thus, high users in the elementary grades are students who took mastery tests more than once per month (5 or more) on the average, while low users took at most one mastery test prior to CAT testing. Table IV-8 shows the percent of students at each grade falling into these Levels of Use categories. While the classification rule adopted does not provide an exact split into thirds for each grade, the proportions in Table IV-8 are roughly equivalent and sufficient for subsequent analysis.

Table IV-8

Percent of Students Classified as High, Medium and Low PRI/RS Users at Each Grade Prior to CAT Testing

	Level of PRI/RS Use		
	Low	Medium	High
1	28	35	37
2	28	31	41
3	29	35	36
4	25	36	39
5	29	34	37
6	35	36	29
7	33	35	32
8	32	30	38

Levels of PRI Use for SES and Ethnic Groups

Having determined the operational definitions of high, medium and low PRI Use to be used in subsequent MANCOVAs, it is of interest to check the generalizability of these standards for students of different ethnic origin and socioeconomic status. Lack of generalizability across these groups will not invalidate subsequent analysis. It would simply result in differential representation of these students in the cells of the design shown in Figure IV-1. This question is investigated here as an interest in itself. That is, SPS policymakers may find it informative to know whether different ethnic or SES groups are using the PRI/RS more or less frequently.

To address this question, students were first classified by grade level, ethnic group, and SES level. The number of mastery tests administered prior to Spring CAT testing was the basis for comparison of these groups. A three-factor analysis of variance (ANOVA) was conducted on this index. The design and analysis is very similar to the previously described ANCOVA, except that there is no need to use a prior measure to statistically equate the student groups before conducting the comparison of interest.

The results of this ANOVA are presented in Table IV-9. Effects whose F-ratios have a p-value (probability of error) less than the conventional .05 will be interpreted here as statistically significant. Tests of four of the seven design effects meet this criterion.

The Grade main effect reaffirms the conclusion already reached--that different grade levels used the PRI with different frequency. The mean frequencies of mastery test use are presented in Table IV-10. As previously noted, the most obvious discrepancy is the decreased use beginning in grade 6, but more markedly in grades 7 and 8. This has been considered in the previously discussed definitions of Levels of PRI Use.

Table IV-9

ANOVA on Number of Mastery Tests Submitted
by Students at Varying Grades, SES and Ethnic Groups

Effect	df	MS	F	p
Grade (G)	7	6695.01	144.97	.000
Ethnic Group (E)	4	73.94	1.60	.171
SES Level (S)	1	386.06	8.36	.004
G X E	28	34.10	.74	.839
G X S	7	153.81	3.33	.002
E X S	4	145.35	3.15	.014
G X E X S	28	72.59	1.57	.128
Residual		46.18		

df = Degrees of Freedom
MS = Hypothesis Mean Square
F = F-ratio
p = p-value (significance level) of F-ratio

Table IV-10

Mean Number of Mastery Tests Submitted
for Significant ANOVA Effects

Grade:

	Gr. 1	Gr. 2	Gr. 3	Gr. 4	Gr. 5	Gr. 6	Gr. 7	Gr. 8
Mean	5.22	5.56	4.98	5.66	5.31	3.98	2.09	2.59
N	3620	3102	3009	3094	3136	3138	3507	3670

SES:

	Low SES	High SES
Mean	4.29	4.42
N	11,833	14,443

Grade & SES:

	Gr. 1	Gr. 2	Gr. 3	Gr. 4	Gr. 5	Gr. 6	Gr. 7	Gr. 8
Low SES								
Mean	4.78	5.11	4.67	5.55	5.44	4.02	1.84	2.51
N	1703	1522	1445	1502	1499	1420	1446	1296
High SES								
Mean	5.61	5.99	5.27	5.76	5.19	3.94	2.26	2.64
N	1917	1580	1564	1592	1637	1718	2061	2374

Ethnic & SES:

	Asian	Black	Hispanic	N. American	White
Low SES					
Mean	3.91	4.30	4.46	4.57	4.49
N	2636	4537	739	570	3351
High SES					
Mean	4.57	4.02	3.97	4.00	4.50
N	1808	1860	520	274	9981

The SES main effect is also statistically significant, and the average Levels of Use for high and low SES students are presented in Table IV-10. High SES students are shown as using the PRI more frequently than low SES students, although the discrepancy is very slight (4.46 vs 4.29 mastery tests submitted). This is another instance where a difference achieves statistical significance, but is probably not educationally significant.

The SES main effect just described must be modified somewhat in the presence of significant interaction effects between the SES factor and the Grade and Ethnic group factors. The Grade by SES interaction evidences an interesting pattern, as described by the means in Table IV-10. High SES students exhibit greater PRI Use than low SES students in the early grades with discrepancies larger than that reported for the SES main effect. This gap gradually declines through grade 4, reverses in grades 5 and 6, and reverses again in grades 7 and 8. Again, some of these differences at the higher grades may not be terribly significant in a practical sense. The authors offer the interpretation that high SES students used the PRI more than low SES students in the early grades, with this discrepancy gradually narrowing and essentially disappearing at the higher grades.

The relationship between SES and student achievement is well documented in volumes of educational research. High SES students typically achieve more highly than low SES students. This is particularly true in the early grades, before remediation or compensatory programs have been implemented to help narrow this gap. With this, the fact that high SES students took more PRI mastery tests than their low SES peers should not be surprising. Acquiring the skills more quickly, they moved through the PRI/RS at a faster rate. This finding in the SPS data, however, has some significance to the current

evaluation. It may suggest that teachers in the early grades were able to implement the system in a more individualized manner, allowing students to work through the objectives at different rates.

The Ethnic by SES interaction is statistically significant, thus adding another moderating influence to the SES main effect reported above. The means in Table IV-10 yield another interesting pattern. Only Asian students evidence the SES main effect pattern, i.e., that high SES students use the PRI more than low SES students. Blacks, Hispanics, and Native Americans all exhibit the opposite pattern--low SES students using the PRI more than high SES students. Interestingly, there is no SES difference for White students. High and low SES White students use the PRI equally often.

The lack of statistical significance in other ANOVA effects in this analysis is of interest in itself. The absence of an Ethnic group main effect indicates that the PRI was used with equal frequency among the various ethnic groups. This is true across all grade levels, as noted by the absence of a Grade by Ethnic interaction. This is a highly positive finding in relation to concerns for equity or bias in PRI implementation.

Quality of PRI Use

In theory, mastery tests are to be administered when a course of study has been completed, or when the teacher judges that a student has sufficiently mastered the skill in question that a test can be given to confirm that mastery. The measure of PRI Use designed to address the primary evaluation question of the study is a simple tally of the number of times a student took a mastery test. As an index of PRI implementation, it is clearly superior to a count of the number of objectives mastered by each student. Yet, as with any statistic, there are still limitations in the PRI Use measure as an index of appropriate use of the system. If high use of the PRI means ill-advised administrations of the tests just to adhere to an implicit demand for frequent testing, this does not suggest proper implementation of the PRI/RS. SPS district staff alerted us to this possibility during our interviews with them in the Design phase of the study. Classroom teachers echoed this concern from their perspective during the Teacher Interview phase of the study (detailed in a subsequent chapter of this report).

The current evaluation is concerned with the quality, as well as quantity, of PRI Use, and additional analyses of the SPS data base were conducted to address this issue. For those students who had taken at least one mastery test prior to Spring CAT testing, another index of PRI Use was calculated, one which is more sensitive to the appropriateness of mastery testing. It combines the two measures of mastery testing discussed above by dividing the number of objectives mastered by the number of mastery tests administered. This ratio is one of successful mastery test use. Students with very low "mastery ratios" have failed most of the mastery tests they have taken. These will include those students who have been unwisely tested too often. Students with high mastery ratios have passed most or all of their mastery tests. This

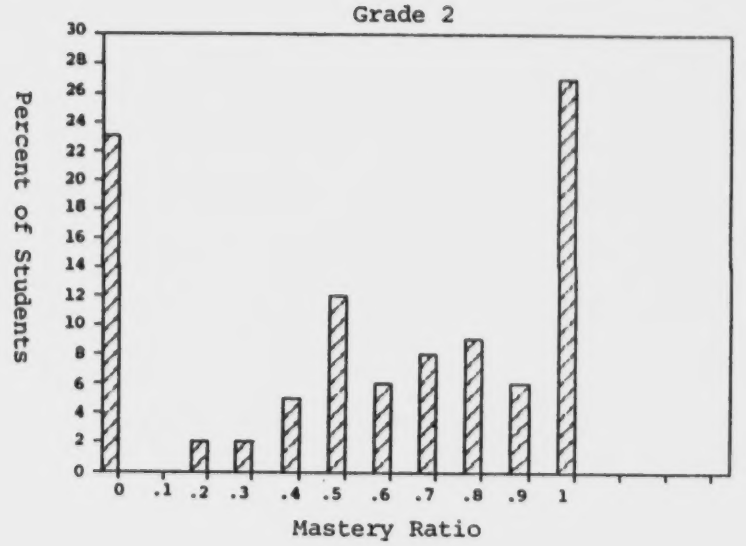
will include students who have not been tested that often, but for whom tests were given at appropriate times, in terms of their understanding of the instructional material.

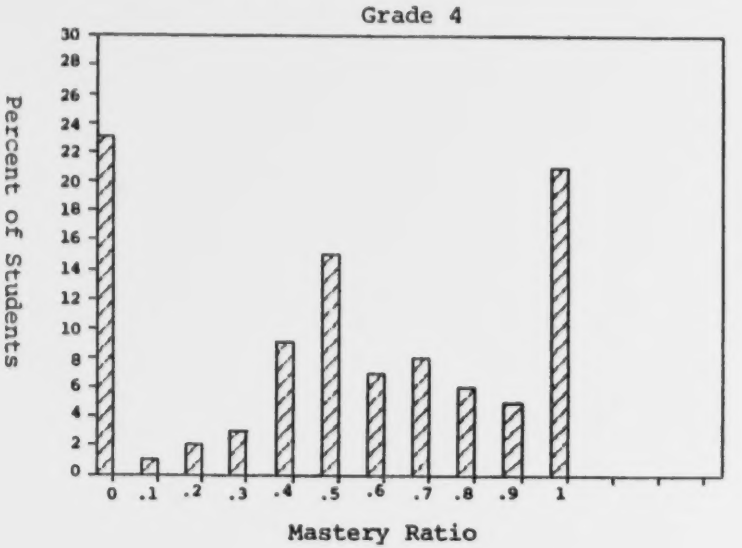
While not a perfect measure of quality of PRI Use, it represents the important aspect of readiness for mastery testing which is critical in the recommended procedures for effective use of the PRI/RS. It is also congruent with one of the highest known correlates of student achievement in current educational research--the proportion of Academic Learning Time in a student's school day. This index is not merely engaged time on task, but one which specifies successful instructional activities in that time.

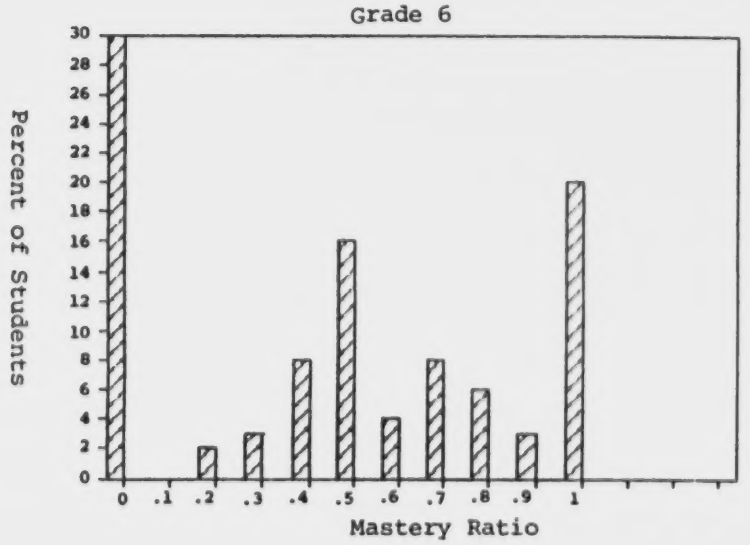
Descriptive statistics for this Mastery Ratio index are given by grade level, before and after Spring CAT testing in Table IV-11. Prior to Spring CAT testing, average success ratios are approximately 50-60 percent through grade 6 and less than 40 percent in grades 7 and 8. In other words, the average elementary student passed a little more than half of the mastery tests he/she took prior to CAT testing; students in middle schools had a lower success rate. After the CAT testing, these Mastery Ratios increase at every grade. At a group level, this does not suggest ill-advised or hasty testing. A typical distribution of success ratios is shown in Figure IV-3. As shown, there are peaks at 0 and 1.00, with a fairly even distribution between these values.

The wide variation in this measure of quality of PRI Use, suggested an additional factor of interest to the authors. Students can be classified according to their success rates in the mastery testing, and their CAT scores compared in much the same way as the PRI Use groupings. For this purpose, and based upon the frequency distributions exemplified in Figure IV-3, the authors define four levels of Mastery Ratio:

Figure IV-3
Illustrative Frequency Distributions of Students'
Mastery Ratio prior to CAT Testing







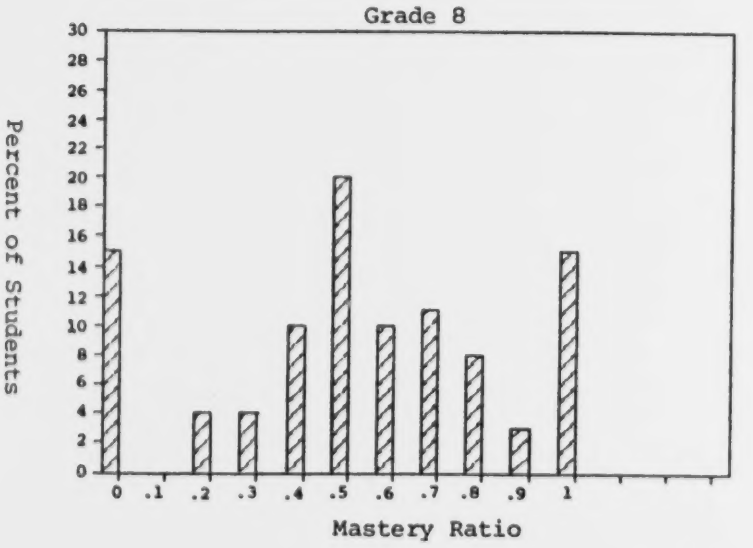


Table IV-11

Mean (M) and Standard Deviation(s) of Mastery Ratios
by Grade Level, Before and After Spring Testing

	Before Spring Testing		After Spring Testing	
	M	S	M	S
1	.56	.38	.59	.26
2	.57	.38	.68	.24
3	.57	.37	.69	.27
4	.51	.36	.67	.28
5	.48	.36	.60	.26
6	.47	.37	.57	.29
7	.37	.39	.54	.27
8	.37	.38	.53	.24

No Mastery - All students whose Mastery Ratio is zero.

At or Below 50% Mastery - Students whose mastery Ratio is greater than zero, but not greater than 50 percent.

Above 50% Mastery - Students whose Mastery Ratio is greater than 50 percent but less than 100 percent.

Full Mastery - Students whose Mastery Ratio is 100 percent.

This factor will be combined with Ethnic groups and SES in a design like that used for the MANCOVA previously described (see Figure IV-1), and the full analysis replicated. It allows another look at the effectiveness of PRI Use, one which may be more sensitive to the quality of mastery testing.

MANCOVA: PRI Use

As previously described, multivariate analyses of covariance (MANCOVA) were conducted on the three-factor design displayed in Figure IV-1, featuring representation of three PRI Use groups, five ethnic groups, and two SES levels. Effects on CAT Reading, Math, and Language Arts scores were analyzed simultaneously.

Results of the three-factor MANCOVAs conducted on the grades 2 through 8 student populations in SPS are given in Tables IV-12 through IV-18. Analysis of 1984 grade 1 data could not be conducted in this way, because previous achievement data (i.e., 1983 kindergarten test results) were not available. In addition, at grade 2, only Reading and Math CAT scores were analyzed, because grade 1 scores in Language Arts were not available.

MANCOVA results for each grade are interpreted using the following guidelines. First, only the four design effects involving PRI Use, as previously described, will be interpreted. This excludes the Ethnic and SES

main effects and their interaction with each other. These relations are already well known to the district, as documented in SPS district-wide assessment reports of the last several years. For the four effects of interest in this study, the multivariate F-ratio will be the first consideration. If it reaches the conventional standard for statistical significance--probability of inferential error less than five percent (i.e., p-value less than .05)--further analysis is interpreted. This consists of univariate F-ratios for the specific subject area test scores, and the discriminant function which identifies the weighted combination of subject areas which maximizes the design effect under study. Finally, when an effect has proven to be statistically significant, and when the nature of the outcome variable is understood, the means of the PRI Use groups will be displayed to interpret the effect in terms of high, medium and low PRI Use.

The MANCOVA results at grade 2 are presented in Table IV-12. The main effect of PRI Use is highly statistically significant (multivariate F=3.71, $p < .005$), and it is clearly evidenced in reading (univariate F=4.54, $p < .011$) and not math (univariate F=14, $p < .319$). Mean CAT Reading scores for high, medium, and low PRI Use groups, adjusted for prior achievement differences, are shown in Table IV-19 in Normal Curve Equivalent (NCE) units. They show the trend which reflects very positively on the effects of PRI Use at this grade level. That is, high PRI users score highest on the CAT Reading test (mean NCE=60.94), medium PRI users next (mean NCE=58.40), and low PRI users lowest (mean NCE=57.39). The absence of any significant interaction effects between PRI Use and the other factors indicates that the highly positive effects of PRI use at this grade level are essentially the same for all ethnic groups and SES levels.

The outcome at other grade levels is not nearly so positive and clear cut. Rather than step through each grade level's results, the authors summarize the MANCOVA results for the four effects of interest in Figure IV-4. Statistically significant effects are noted with an (X) in the grade levels and subject areas in which they are found. Detailed statistical results corresponding to this summary are found in Tables IV-12 through IV-18.

The summary in Figure IV-4 indicates that statistically significant PRI Use main effects are found at four of the seven grades--2, 3, 5 and 7--on some combination of subject areas. A closer look through Tables IV-12 to 18 shows that the results at grade 6 were very close to the required level of significance ($p < .061$), but grades 4 and 8 were not. Thus, there is a mixed result of PRI effectiveness across grade levels. The mixture continues in looking at the subject areas that manifest these effects at the selected grade levels. It is not always Reading that evidences significant differences amount PRI Use groups. At grades 3 and 5 it is Language Arts, while at grades 2 and 7 it is Reading. Math results are also significant at grade 5, though less prominently than Language Arts.

To complete the interpretation of the PRI Use main effects across grade levels, the adjusted means of the significant subject areas are highlighted (shaded area) in Table IV-19. In Reading, as already discussed, the results at grade 2 show the very positive trend. At grade 7, while the differences are statistically significant, the adjusted means show the inverse of the hypothesized trend--the high PRI Use group is the lowest achieving (mean NCE=59.88) while the low PRI use group is the highest achieving (mean NCE=63.13). The magnitude of the difference shown is comparable to that previously described for grade 2--about 3.5 NCEs. Interestingly, if we "bent"

Table IV-12

Multivariate Analysis of Covariance[1]
for Ethnic by SES by PRI Use Design
Grade 2

Effect	df	MS	F	p	b
ETHNIC (E)	8		13.61	.000	
Reading	4	1855.36	11.46	.000	.18
Math	4	4166.30	23.55	.000	.90
SES (S)	2		8.87	.000	
Reading	1	1810.88	11.18	.001	.46
Math	1	2641.77	14.94	.000	.70
PRI USE (P)	4		3.71	.005	
Reading	2	734.37	4.54	.011	1.13
Math	2	202.16	1.14	.319	- .67
E x S	8		2.02	.040	
Reading	4	398.25	2.46	.044	.58
Math	4	437.67	2.47	.043	.58
E x P	16		1.35	.158	
Reading	8	109.83	.68	.711	
Math	8	278.61	1.58	.127	
S x P	4		.40	.807	
Reading	2	110.72	.68	.505	
Math	2	11.73	.07	.936	
E x S x P	16		.84	.645	
Reading	8	123.56	.76	.635	
Math	8	210.57	1.19	.301	

[1] Multivariate test statistics are given on line where the effect is listed. Univariate statistics for Reading, Math and Language Arts are given immediately below.

df = Degrees of Freedom
MS = Hypothesis Mean Square
F = F-ratio
p = p-value (significance level) of F-ratio
b = Standardized coefficient in statistically significant ($p < .05$) discriminant functions

Figure IV-4

Summary of MANCOVA Results
PRI Use Design, Grades 2-8

Effect	Grade	Subject Area			
		Composite[1]	Reading	Language Arts	Math
PRI Use (P) Main Effect	2	X	X		
	3	X		X	
	4				
	5	X		X	X
	6				
	7	X	X		
	8				
P x Ethnic Interaction	2				
	3				
	4				
	5	X		X	
	6				
	7				
	8				
P x SES Interaction	2				
	3				
	4				
	5				
	6				
	7	X		X	
	8				
P x E x SES Interaction	2				
	3				
	4				
	5				
	6				
	7	X		X	
	8				

X Indicates statistically significant F-ratio ($p < .05$)

1 "Composite" is the multivariate combination of all three subject areas.

Table IV-13

Multivariate Analysis of Covariance[1]
for Ethnic by SES by PRI Use Design
Grade 3

Effect	df	MS	F	p	b
ETHNIC (E)	12		7.12	.000	
Reading	4	904.35	6.27	.000	.48
Math	4	2103.27	12.66	.000	-1.15
Language	4	899.54	5.17	.000	- .04
SES (S)	3		4.46	.004	
Reading	1	1769.11	12.27	.000	.82
Math	1	549.96	3.31	.069	- .11
Language	1	1329.25	7.65	.006	.36
PRI USE (P)	6		2.55	.018	
Reading	2	102.35	.71	.492	- .23
Math	2	167.40	1.01	.365	- .21
Language	2	970.34	5.58	.004	1.18
E x S	12		1.63	.075	
Reading	4	361.68	2.51	.040	
Math	4	296.39	1.78	.129	
Language	4	196.19	1.13	.341	
E x P	24		1.20	.226	
Reading	8	122.44	.85	.559	
Math	8	103.06	.62	.761	
Language	8	265.31	1.53	.143	
S x P	6		.96	.449	
Reading	2	149.67	1.04	.354	
Math	2	395.67	2.38	.093	
Language	2	182.62	1.05	.350	
E x S x P	24		.80	.744	
Reading	8	186.07	1.29	.244	
Math	8	99.34	.60	.780	
Language	8	216.08	1.24	.270	

[1] Multivariate test statistics are given on line where the effect is listed. Univariate statistics for Reading, Math and Language Arts are given immediately below.

df = Degrees of Freedom
MS = Hypothesis Mean Square
F = F-ratio
p = p-value (significance level) of F-ratio
b = Standardized coefficient in statistically significant ($p < .05$) discriminant functions

Table IV-14

Multivariate Analysis of Covariance[1]
for Ethnic by SES by PRI Use Design
Grade 4

Effect	df	MS	F	p	b
ETHNIC (E)	12		17.15	.000	
Reading	4	2027.04	16.25	.000	.48
Math	4	4264.09	30.46	.000	-1.00
Language	4	2495.78	15.91	.000	- .21
SES (S)	3		4.78	.003	
Reading	1	1769.11	12.27	.000	.94
Math	1	186.38	1.33	.249	- .23
Language	1	812.34	5.18	.023	.27
PRI USE (P)	6		.94	.467	
Reading	2	27.53	.22	.802	
Math	2	282.22	2.02	.134	
Language	2	56.81	.36	.696	
E x S	12		2.21	.009	
Reading	4	277.92	2.23	.064	.11
Math	4	563.44	4.02	.003	-1.00
Language	4	294.16	1.88	.112	- .07
E x P	24		.85	.677	
Reading	8	154.82	1.24	.271	
Math	8	148.14	1.06	.390	
Language	8	108.08	.69	.702	
S x P	6		.90	.497	
Reading	2	77.06	.62	.539	
Math	2	179.15	1.28	.278	
Language	2	168.49	1.07	.342	
E x S x P	24		.75	.806	
Reading	8	151.64	1.22	.285	
Math	8	177.31	1.27	.257	
Language	8	74.01	.47	.877	

[1] Multivariate test statistics are given on line where the effect is listed. Univariate statistics for Reading, Math and Language Arts are given immediately below.

df = Degrees of Freedom
MS = Hypothesis Mean Square
F = F-ratio
p = p-value (significance level) of F-ratio
b = Standardized coefficient in statistically significant ($p < .05$) discriminant functions

Table IV-15

Multivariate Analysis of Covariance[1]
for Ethnic by SES by PRI Use Design
Grade 5

Effect	df	MS	F	p	b
ETHNIC (E)	12		10.13	.000	
Reading	4	986.38	9.97	.000	- .04
Math	4	2837.40	21.45	.000	1.00
Language	4	852.19	6.37	.000	.04
SES (S)	3		9.02	.000	
Reading	1	2262.11	22.87	.000	- .72
Math	1	980.95	7.41	.007	- .10
Language	1	1947.68	14.55	.000	- .39
PRI USE (P)	6		6.05	.000	
Reading	2	5.24	.05	.948	- .49
Math	2	936.69	7.08	.001	- .43
Language	2	1750.72	13.08	.000	.88
E x S	12		.75	.707	
Reading	4	29.61	.30	.878	
Math	4	73.37	.55	.696	
Language	4	151.56	1.13	.339	
E x P	24		1.80	.009	
Reading	8	186.83	1.89	.058	- .47
Math	8	172.25	1.30	.238	- .34
Language	8	291.83	2.18	.026	1.15
S x P	6		.24	.962	
Reading	2	11.02	.11	.895	
Math	2	38.84	.29	.746	
Language	2	22.64	.17	.844	
E x S x P	24		1.20	.231	
Reading	8	137.10	1.39	.197	
Math	8	69.33	.52	.839	
Language	8	179.34	1.34	.219	

[1] Multivariate test statistics are given on line where the effect is listed. Univariate statistics for Reading, Math and Language Arts are given immediately below.

df = Degrees of Freedom
MS = Hypothesis Mean Square
F = F-ratio
p = p-value (significance level) of F-ratio
b = Standardized coefficient in statistically significant ($p < .05$) discriminant functions

Table IV-16

Multivariate Analysis of Covariance[1]
for Ethnic by SES by PRI Use Design
Grade 6

Effect	df	MS	F	p	b
ETHNIC (E)	12		7.82	.000	
Reading	4	443.83	4.50	.001	- .26
Math	4	1812.11	16.60	.000	.96
Language	4	737.20	6.39	.000	.22
SES (S)	3		4.14	.006	
Reading	1	1062.75	10.79	.001	- .78
Math	1	980.95	7.41	.007	- .10
Language	1	457.77	3.97	.046	- .21
PRI USE (P)	6		2.01	.061	
Reading	2	288.35	2.93	.054	
Math	2	234.78	2.15	.117	
Language	2	48.91	.42	.655	
E x S	12		2.28	.007	
Reading	4	65.81	.67	.614	.11
Math	4	483.93	4.43	.001	-1.08
Language	4	108.87	.94	.438	.46
E x P	24		1.17	.257	
Reading	8	94.28	.96	.468	
Math	8	116.47	1.07	.383	
Language	8	204.79	1.78	.077	
S x P	6		1.82	.091	
Reading	2	173.28	1.76	.173	
Math	2	451.18	4.13	.016	
Language	2	243.53	2.11	.121	
E x S x P	24		.72	.840	
Reading	8	101.29	1.03	.412	
Math	8	91.45	.84	.569	
Language	8	69.97	.61	.773	

[1] Multivariate test statistics are given on line where the effect is listed. Univariate statistics for Reading, Math and Language Arts are given immediately below.

df = Degrees of Freedom
MS = Hypothesis Mean Square
F = F-ratio
p = p-value (significance level) of F-ratio
b = Standardized coefficient in statistically significant ($p < .05$) discriminant functions

Table IV-17

Multivariate Analysis of Covariance[1]
for Ethnic by SES by PRI Use Design
Grade 7

Effect	df	MS	F	p	b
ETHNIC (E)	12		9.96	.000	
Reading	4	708.41	6.59	.000	.39
Math	4	1933.19	19.59	.000	-1.06
Language	4	276.54	2.66	.031	.21
SES (S)	3		6.46	.000	
Reading	1	2044.69	19.02	.000	.96
Math	1	96.56	.98	.323	- .08
Language	1	381.57	3.67	.056	.14
PRI USE (P)	6		4.54	.000	
Reading	2	1051.95	9.78	.000	- .93
Math	2	275.05	2.79	.062	- .33
Language	2	283.97	2.73	.065	.23
E x S	12		1.80	.042	
Reading	4	136.64	1.27	.279	- .29
Math	4	369.11	3.74	.005	.98
Language	4	80.56	.77	.541	.19
E x P	24		1.36	.115	
Reading	8	60.60	.56	.808	
Math	8	174.04	1.76	.080	
Language	8	131.70	1.27	.256	
S x P	6		2.50	.020	
Reading	2	224.52	2.09	.124	.25
Math	2	74.64	.76	.470	- .07
Language	2	718.10	6.91	.001	.91
E x S x P	24		1.89	.005	
Reading	8	195.10	1.81	.070	- .47
Math	8	115.21	1.17	.315	- .28
Language	8	236.53	2.28	.020	1.07

[1] Multivariate test statistics are given on line where the effect is listed. Univariate statistics for Reading, Math and Language Arts are given immediately below.

df = Degrees of Freedom
MS = Hypothesis Mean Square
F = F-ratio
p = p-value (significance level) of F-ratio
b = Standardized coefficient in statistically significant ($p < .05$) discriminant functions

Table IV-18

Multivariate Analysis of Covariance[1]
for Ethnic by SES by PRI Use Design
Grade 8

Effect	df	MS	F	p	b
ETHNIC (E)	12		10.93	.000	
Reading	4	368.07	4.55	.001	.23
Math	4	1794.38	23.29	.000	-1.01
Language	4	794.15	7.56	.000	- .03
SES (S)	3		3.28	.020	
Reading	1	394.22	4.87	.027	.57
Math	1	8.43	.11	.741	- .41
Language	1	618.40	5.89	.015	.71
PRI USE (P)	6		.65	.694	
Reading	2	81.33	1.01	.366	
Math	2	23.46	.30	.738	
Language	2	76.20	.73	.484	
E x S	12		1.70	.059	
Reading	4	190.36	2.35	.052	
Math	4	102.49	1.33	.256	
Language	4	60.49	.58	.680	
E x P	24		1.18	.252	
Reading	8	109.12	1.35	.214	
Math	8	115.41	1.50	.153	
Language	8	61.37	.58	.791	
S x P	6		1.00	.422	
Reading	2	131.57	1.63	.197	
Math	2	37.84	.49	.612	
Language	2	46.93	.45	.640	
E x S x P	24		1.20	.226	
Reading	8	160.78	1.99	.044	
Math	8	82.76	1.07	.378	
Language	8	54.23	.52	.845	

[1] Multivariate test statistics are given on line where the effect is listed. Univariate statistics for Reading, Math and Language Arts are given immediately below.

df = Degrees of Freedom
MS = Hypothesis Mean Square
F = F-ratio
p = p-value (significance level) of F-ratio
b = Standardized coefficient in statistically significant ($p < .05$) discriminant functions

Table IV-19

Adjusted Mean NCEs for PRI Use Groups
Grades 2-8

Grade:	2	3	4	5	6	7	8
PRI USE							
Reading							
Low	57.39	57.85	55.18	59.21	61.84	63.13	61.39
Medium	58.40	56.83	55.61	59.26	61.61	61.02	60.66
High	60.94	56.92	54.32	59.54	60.92	59.88	59.76
Math							
Low	61.37	59.68	56.56	59.18	62.08	61.95	61.31
Medium	61.80	58.89	56.48	61.32	62.06	60.83	63.08
High	63.91	60.12	56.00	63.47	61.39	61.24	61.72
Language Arts							
Low		57.07	56.59	60.66	60.62	59.04	60.43
Medium		57.09	57.66	59.92	61.93	59.11	61.00
High		58.36	57.46	63.61	61.32	58.88	59.86

☐ Statistically significant differences ($p < .05$)

Table IV-20

Adjusted Mean NCEs for Significant Interaction Effects
PRI Use Design, Grades 5 and 7 Language Arts

Grade 5 P x Ethnic Interaction	Low	Medium	High
Asian	60.67	62.42	64.70
Black	58.54	60.24	62.01
Hispanic	55.16	62.07	63.18
Native American	67.20	54.06	63.81
White	61.71	60.81	64.33
Grade 7 **P x SES Interaction**			
Low SES	57.78	58.73	59.18
High SES	59.97	59.50	58.90

our statistical significance criterion for interpretation, and elected to consider the PRI Use effect at grade 6 significant, the direction of this effect evidences the same negative trend as in grade 7, though not as large. In Language Arts, PRI Use effects are significant at grades 3 and 5. Here, the direction of the differences is a positive comment on PRI effectiveness. The high PRI Use group is the highest achieving, differing from medium and low PRI Use groups, which do not differ significantly from each other.

There is a conspicuous absence of significant interaction effects across the grade levels, indicating that the effects of PRI Use rarely vary by ethnic group or socioeconomic status. This is a highly positive finding in terms of justice or equity issues. That is, the PRI/RS shows little evidence of containing any ethnic or SES bias in its attempts to enhance reading skills. The exceptions occur at grades 5 and 7, but only in Language Arts. Adjusted means which facilitate the interpretation of these interactions are presented in Table IV-20. At grade 5, there is a significant Ethnic by PRI Use interaction in Language Arts. In other words, the positive effects of PRI Use noted above vary by ethnic group. Asian, Black and Hispanic students all evidence the highly positive effect of high PRI users exhibiting the highest achievement and low PRI users showing the lowest. White students also show high PRI users with the highest scores in Language Arts, but medium and low users are virtually the same. Native American students are the only ethnic group which departs drastically from this trend. Medium PRI users are dramatically lower in achievement than high or low users. At grade 7, there is both a significant PRI Use by SES interaction and a significant higher order interaction involving PRI Use, Ethnic Origin, and SES. Both are manifested in Language Arts, and the adjusted means for the PRI Use by SES

interaction are shown in Table IV-20. The PRI Use by SES interaction shows that low SES students evidence the hypothesized trend in achievement for varying levels of PRI use, while high SES students show the inverse. The significant triple interaction indicates that even this pattern is not consistent across ethnic groups.

These results at specific grade levels are interesting in themselves, but lack generality in the context of such a broadly based implementation as SPS experienced with the PRI/RS. Varying levels of use evidenced significant effects at grades 2, 3 and 5; but, mysteriously, not grade 4. Some grade levels showed that the PRI affects Reading scores, while some showed the effects in Language Arts. There was no consistent trend with increasing or decreasing grades. Furthermore, there was an instance when the highest PRI users exhibited the lowest achievement (Reading, grade 7).

MANCOVA: Mastery Ratio

In earlier presentations in this report of the frequency of PRI Use, it was noted that the distributions were highly skewed--over half of the students were concentrated below the mean with a non-trivial proportion of students showing a surprisingly high number of times tested. This reinforced SPS district and building staff concerns for quality of PRI/RS implementation, and led the authors to formulate an alternative index of PRI Use. The Mastery Ratio is the proportion of mastery tests on which the student successfully demonstrated mastery.

MANCOVAs were performed in grades 2 through 8 using this new index of use, Mastery Ratio, in place of the PRI Use factor in a three-factor design as shown in Figure IV-1. That is, students were classified into one of the four

Figure IV-5

Summary of MANCOVA Results
Mastery Ratio Design, Grades 2-8

Effect	Grade	Subject Area			
		Composite[1]	Reading	Language Arts	Math
Mastery Ratio (M) Main Effect	2	X	X		
	3	X	X	X	
	4	X		X	X
	5	X	X	X	X
	6	X		X	X
	7	X	X		X
	8	X		X	X
M x Ethnic (E) Interaction	2				
	3				
	4				
	5				
	6				
	7				
	8				
M x SES Interaction	2				
	3				
	4				
	5				
	6				
	7	X	X	X	
	8				
M x E x SES Interaction	2				
	3				
	4				
	5				
	6				
	7				
	8				

X Indicates statistically significant F-ratio ($p < .05$)

1 "Composite" is the multivariate combination of all three subject areas.

Table IV-21

Multivariate Analysis of Covariance[1]
for Ethnic by SES by Mastery Ratio Design
Grade 2

Effect	df	MS	F	p	b
ETHNIC (E)	8		14.03	.000	
Reading	4	1916.49	12.05	.000	.21
Math	4	4239.91	23.96	.000	.88
SES (S)	2		9.44	.000	
Reading	1	1897.79	11.93	.001	.46
Math	1	2797.15	15.81	.001	.69
MASTRATE (M)	6		6.33	.000	
Reading	3	1810.81	11.39	.000	1.07
Math	3	421.95	2.38	.067	- .17
E x S	8		2.12	.031	
Reading	4	421.20	2.65	.032	.62
Math	4	446.66	2.52	.039	.54
E x M	24		1.22	.214	
Reading	12	256.17	1.61	.082	
Math	12	157.69	.89	.555	
S x M	6		.36	.904	
Reading	3	90.85	.57	.634	
Math	3	6.23	.04	.991	
E x S x M	24		.97	.505	
Reading	12	130.00	.82	.633	
Math	12	234.40	1.32	.197	

[1] Multivariate test statistics are given on line where the effect is listed. Univariate statistics for Reading, Math and Language Arts are given immediately below.

df = Degrees of Freedom
MS = Hypothesis Mean Square
F = F-ratio
p = p-value (significance level) of F-ratio
b = Standardized coefficient in statistically significant ($p < .05$) discriminant functions

Table IV-22

Multivariate Analysis of Covariance[1]
for Ethnic by SES by Mastery Ratio Design
Grade 3

Effect	df	MS	F	p	b
ETHNIC (E)	12		7.28	.000	
Reading	4	966.64	6.73	.001	.47
Math	4	2131.72	12.83	.000	-1.15
Language	4	937.72	5.43	.000	- .05
SES (S)	3		4.70	.003	
Reading	1	1863.54	12.98	.000	.83
Math	1	552.11	3.32	.069	-1.22
Language	1	1361.92	7.89	.005	.35
MASTRATE (M)	9		3.70	.000	
Reading	3	781.52	5.44	.001	.36
Math	3	304.65	1.83	.139	- .25
Language	3	1617.58	9.37	.000	.88
E x S	12		1.70	.061	
Reading	4	365.72	2.55	.038	
Math	4	310.31	1.87	.114	
Language	4	186.83	1.08	.364	
E x M	36		.89	.651	
Reading	12	137.20	.96	.490	
Math	12	142.88	.86	.588	
Language	12	166.48	.96	.481	
S x M	9		.89	.536	
Reading	3	41.21	.29	.835	
Math	3	125.63	.76	.519	
Language	3	76.30	.41	.723	
E x S x M	36		.95	.552	
Reading	12	113.67	.79	.660	
Math	12	102.40	.62	.830	
Language	12	248.17	1.44	.142	

[1] Multivariate test statistics are given on line where the effect is listed. Univariate statistics for Reading, Math and Language Arts are given immediately below.

df = Degrees of Freedom
MS = Hypothesis Mean Square
F = F-ratio
p = p-value (significance level) of F-ratio
b = Standardized coefficient in statistically significant ($p < .05$) discriminant functions

Table IV-23

Multivariate Analysis of Covariance[1]
for Ethnic by SES by Mastery Ratio Design
Grade 4

Effect	df	MS	F	p	b
ETHNIC (E)	12		17.32	.000	
Reading	4	1990.74	16.09	.000	.51
Math	4	4243.52	30.38	.000	-1.00
Language	4	2525.89	16.35	.000	- .23
SES (S)	3		4.79	.002	
Reading	1	1663.18	13.44	.000	.94
Math	1	163.90	1.17	.279	- .26
Language	1	800.50	5.18	.023	.27
MASTRATE (M)	9		2.57	.006	
Reading	3	141.35	1.14	.331	- .08
Math	3	477.51	3.42	.017	.05
Language	3	759.56	4.92	.002	1.01
E x S	12		2.30	.006	
Reading	4	277.78	2.24	.062	.14
Math	4	588.05	4.21	.002	-1.00
Language	4	308.52	2.00	.093	- .09
E x M	36		1.32	.093	
Reading	12	170.41	1.38	.170	
Math	12	144.77	1.04	.412	
Language	12	222.49	1.44	.141	
S x M	9		1.41	.179	
Reading	3	292.89	2.37	.069	
Math	3	19.77	.14	.935	
Language	3	137.07	.89	.447	
E x S x M	36		1.21	.186	
Reading	12	188.40	1.52	.109	
Math	12	182.43	1.31	.208	
Language	12	177.11	1.15	.318	

[1] Multivariate test statistics are given on line where the effect is listed. Univariate statistics for Reading, Math and Language Arts are given immediately below.

df = Degrees of Freedom
MS = Hypothesis Mean Square
F = F-ratio
p = p-value (significance level) of F-ratio
b = Standardized coefficient in statistically significant ($p < .05$) discriminant functions

Table IV-24

Multivariate Analysis of Covariance[1]
for Ethnic by SES by Mastery Ratio Design
Grade 5

Effect	df	MS	F	p	b
ETHNIC (E)	12		10.27	.000	
Reading	4	999.61	10.15	.000	.02
Math	4	2867.35	21.80	.000	- .99
Language	4	870.64	6.52	.000	- .05
SES (S)	3		9.28	.000	
Reading	1	2293.91	23.30	.000	- .71
Math	1	1001.74	7.62	.006	- .11
Language	1	1995.02	14.95	.000	- .40
MASTRATE (M)	9		4.75	.000	
Reading	3	525.05	5.33	.001	- .21
Math	3	914.63	6.95	.000	- .34
Language	3	1589.36	11.91	.000	- .69
E x S	12		.71	.744	
Reading	4	33.29	.34	.852	
Math	4	69.19	.53	.717	
Language	4	136.85	1.03	.393	
E x M	36		1.22	.176	
Reading	12	84.37	.86	.591	
Math	12	183.84	1.40	.159	
Language	12	202.67	1.52	.110	
S x M	9		.87	.555	
Reading	3	93.56	.95	.415	
Math	3	239.26	1.82	.142	
Language	3	79.72	.60	.617	
E x S x M	36		1.02	.443	
Reading	12	141.43	1.44	.142	
Math	12	105.66	.80	.647	
Language	12	169.32	1.27	.230	

[1] Multivariate test statistics are given on line where the effect is listed. Univariate statistics for Reading, Math and Language Arts are given immediately below.

df = Degrees of Freedom
MS = Hypothesis Mean Square
F = F-ratio
p = p-value (significance level) of F-ratio
b = Standardized coefficient in statistically significant ($p < .05$) discriminant functions

Table IV-25

Multivariate Analysis of Covariance[1]
for Ethnic by SES by Mastery Ratio Design
Grade 6

Effect	df	MS	F	p	b
ETHNIC (E)	12		7.81	.000	
Reading	4	423.09	4.26	.002	.28
Math	4	1807.57	16.61	.000	- .96
Language	4	744.58	6.46	.000	- .23
SES (S)	3		3.96	.008	
Reading	1	1014.08	10.21	.001	- .77
Math	1	452.74	4.16	.042	- .27
Language	1	459.59	3.99	.046	.23
MASTRATE (M)	9		3.39	.000	
Reading	3	57.09	.57	.632	.47
Math	3	553.62	5.09	.002	- .61
Language	3	594.18	5.15	.001	- .71
E x S	12		2.38	.005	
Reading	4	69.90	.70	.589	.11
Math	4	522.51	4.80	.001	-1.09
Language	4	108.03	.94	.441	.42
E x M	36		.55	.987	
Reading	12	48.85	.49	.921	
Math	12	64.99	.60	.846	
Language	12	84.47	.73	.720	
S x M	9		1.10	.360	
Reading	3	30.04	.30	.824	
Math	3	266.79	2.45	.062	
Language	3	177.06	1.54	.203	
E x S x M	36		.77	.839	
Reading	12	62.46	.63	.819	
Math	12	115.96	1.07	.386	
Language	12	68.27	.59	.850	

[1] Multivariate test statistics are given on line where the effect is listed. Univariate statistics for Reading, Math and Language Arts are given immediately below.

df = Degrees of Freedom
MS = Hypothesis Mean Square
F = F-ratio
p = p-value (significance level) of F-ratio
b = Standardized coefficient in statistically significant ($p < .05$) discriminant functions

Table IV-26

Multivariate Analysis of Covariance[1]
for Ethnic by SES by Mastery Ratio Design
Grade 7

Effect	df	MS	F	p	b
ETHNIC (E)	12		9.83	.000	
Reading	4	672.93	6.26	.000	- .38
Math	4	1924.32	19.55	.000	1.06
Language	4	270.96	2.59	.035	- .21
SES (S)	3		6.24	.000	
Reading	1	1968.23	18.31	.000	.96
Math	1	78.92	.80	.371	- .10
Language	1	372.90	3.57	.059	.16
MASTRATE (M)	9		3.45	.000	
Reading	3	591.40	5.50	.001	- .67
Math	3	573.06	5.82	.001	- .69
Language	3	73.95	.71	.548	.32
E x S	12		1.81	.041	
Reading	4	148.42	1.38	.238	- .21
Math	4	367.45	3.73	.005	.96
Language	4	92.74	.89	.471	.20
E x M	36		.91	.619	
Reading	12	113.59	1.06	.393	
Math	12	84.86	.86	.586	
Language	12	60.43	.58	.861	
S x M	9		2.39	.011	
Reading	3	297.17	2.76	.041	- .41
Math	3	149.10	1.51	.209	.02
Language	3	557.83	5.34	.001	- .80
E x S x M	36		1.42	.048	
Reading	12	133.27	1.24	.249	.39
Math	12	115.22	1.17	.299	- .74
Language	12	168.42	1.61	.082	.79

[1] Multivariate test statistics are given on line where the effect is listed. Univariate statistics for Reading, Math and Language Arts are given immediately below.

df = Degrees of Freedom
MS = Hypothesis Mean Square
F = F-ratio
p = p-value (significance level) of F-ratio
b = Standardized coefficient in statistically significant ($p < .05$) discriminant functions

Table IV-27

Multivariate Analysis of Covariance[1]
for Ethnic by SES by Mastery Ratio Design
Grade 8

Effect	df	MS	F	p	b
ETHNIC (E)	12		10.88	.000	
Reading	4	342.98	4.25	.002	.23
Math	4	1808.72	23.46	.000	-1.01
Language	4	788.65	7.56	.000	- .05
SES (S)	3		3.15	.024	
Reading	1	362.94	4.50	.034	.56
Math	1	9.80	.13	.721	- .41
Language	1	601.40	5.77	.016	.73
MASTRATE (M)	9		2.89	.002	
Reading	3	10.29	.13	.944	- .21
Math	3	289.36	3.75	.011	.47
Language	3	639.34	6.13	.000	.82
E x S	12		1.68	.064	
Reading	4	178.16	2.21	.066	
Math	4	114.20	1.48	.205	
Language	4	52.53	.50	.733	
E x M	36		1.04	.406	
Reading	12	118.38	1.47	.129	
Math	12	77.92	1.01	.436	
Language	12	104.20	1.00	.447	
S x M	9		.41	.933	
Reading	3	44.91	.56	.644	
Math	3	8.41	.11	.955	
Language	3	48.75	.47	.705	
E x S x M	36		1.02	.437	
Reading	12	188.22	2.33	.006	
Math	12	44.87	.58	.858	
Language	12	44.48	.43	.954	

[1] Multivariate test statistics are given on line where the effect is listed. Univariate statistics for Reading, Math and Language Arts are given immediately below.

df = Degrees of Freedom
MS = Hypothesis Mean Square
F = F-ratio
p = p-value (significance level) of F-ratio
b = Standardized coefficient in statistically significant ($p < .05$) discriminant functions

previously defined Mastery Ratio groups (No Mastery to Full Mastery) along with their ethnic group and SES level, and their CAT scores in Reading, Language Arts and Math were analyzed. Results of these analyses are presented in Tables IV-21 through IV-27, and are summarized in Figure IV-5. As shown in the summary, the Mastery Ratio main effect is significant at all grade levels. Reading evidences the effect at grades 2, 3, 5 and 7, while Language Arts appears at grades 3, 4, 5, 6 and 8. Math also shows this effect in grades 4-8, though typically secondary to Reading or Language Arts in magnitude. At all but grade 2, at least two of the subject areas manifests significant differences across Mastery Ratio groups. The standardized discriminant function weights (denoted "b" in Tables IV-21 through IV-27) indicate the relative importance of the subject areas in differentiating among the four Ratio groups. In every case, this is either Reading (grades 2 and 7) or Language Arts (grades 3, 4, 5, 6 and 8).

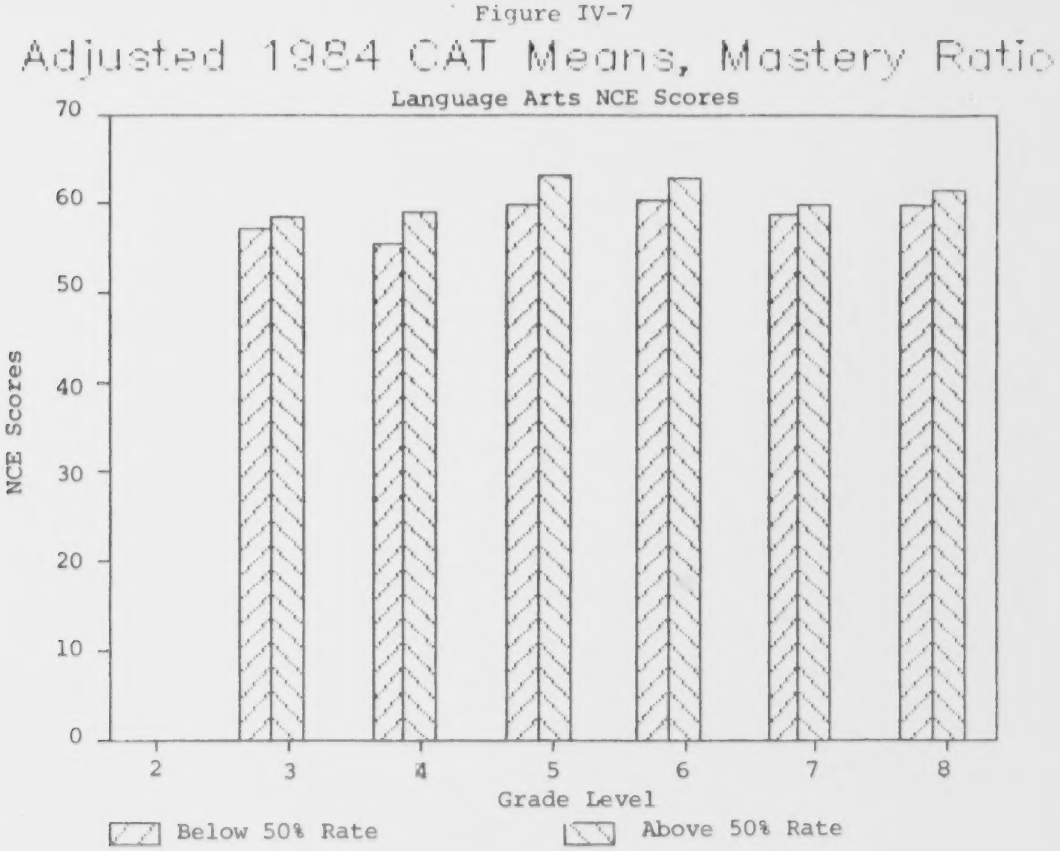
The adjusted means for the four Mastery Ratio groups are presented in Table IV-28. The grade level/subject matter combinations which evidenced significant differences in the MANCOVAs are highlighted. In Reading, the highest achieving group is the 100 percent mastery group at grades 2, 3 and 5. As in the PRI Use analysis, differences are highest at the early grade levels. At grade 2, the difference in Reading achievement between the zero and Full Mastery groups is 6 NCEs. As grade level increases, this discrepancy is on the order of 2 NCEs. In grades 7 and 8, differences are even smaller, with a less consistent trend among the four Ratio groups. In Language Arts, the highest achieving group is either the 50-99 percent mastery group or the 100 percent mastery group at all five of the grade levels where the differences are significant. Unlike Reading, differences in Language Arts are greater at higher grade levels--typically about 4 NCEs through grade 6. One

might begin to infer that PRI effectiveness is evidenced in Reading in the early grades, yielding to Language Arts in the intermediate grades.

The adjusted mean CAT scores shown on Table IV-28 do not always evidence the most positive, monotonic trend in achievement across the four Mastery Ratio groups. That is, the Full Mastery group is not always the highest achieving, the Above 50% next highest, etc. In fact, in many instances, the Above 50% group is the highest achieving. This may reflect limitations in our Mastery Ratio index as we acknowledged previously. Full Mastery students include those who may have been tested only once or twice, mastering the few objectives on which they had been tested. To the extent that such low frequency of use (i.e., quantity) students dominate the Full Mastery group, it is not surprising that this is not always the highest achieving group. The composition of the four Mastery Ratio groups may be of further interest, particularly to those concerned with recommendations for effective teacher implementation of the PRI/RS.

By combining students in the two lowest Mastery Ratio groups, and comparing them against the two highest Mastery Ratio groups, the trends in achievement are most consistent. As shown in Table IV-29, the "Above 50% Mastery" group achieves more highly than the "Below 50% Mastery" group at all grades in Language Arts and in all but grades 6 and 8 in Reading, where the difference is less than one-half an NCE score. The effects of quality PRI Use are most prominent at grade 2 in Reading (6 NCE difference) and grades 4 and 5 in Language Arts (3-4 NCEs). Furthermore, these positive effects essentially disappear by grade 6 in Reading, but are maintained through grade 8 in Language Arts. These trends across grades are depicted graphically in Figure IV-6. Their consistency is in contrast to those previously reported among PRI Use groups, depicted in Figure IV-7.

Figure IV-7


Adjusted 1984 CAT Means, Mastery Ratio
Language Arts NCE Scores
NCE Scores
70
60
50
40
30
20
10
0
2
3
4
5
6
7
8
Grade Level
Below 50% Rate
Above 50% Rate

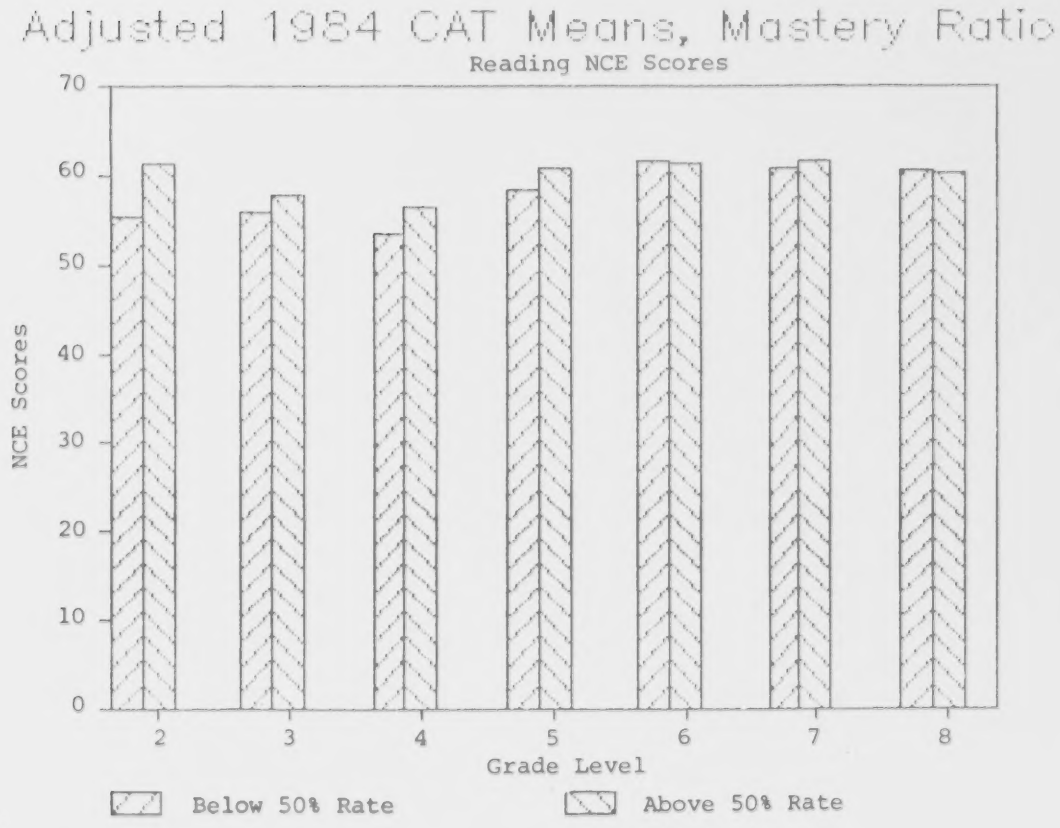
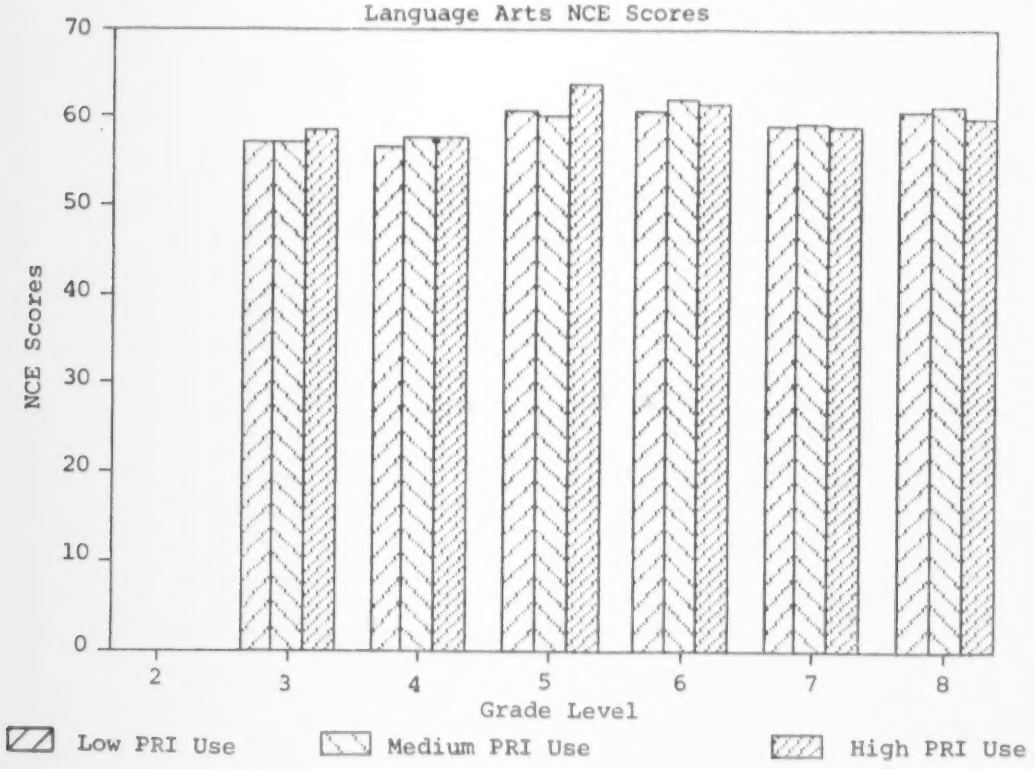

Adjusted 1984 CAT Means, Mastery Ratio
Reading NCE Scores
NCE Scores
70
60
50
40
30
20
10
0
2
3
4
5
6
7
8
Grade Level
Below 50% Rate
Above 50% Rate

Figure IV-8

Adjusted 1984 CAT Means, PRI Use Groups



Adjusted 1984 CAT Means, PRI Use Groups

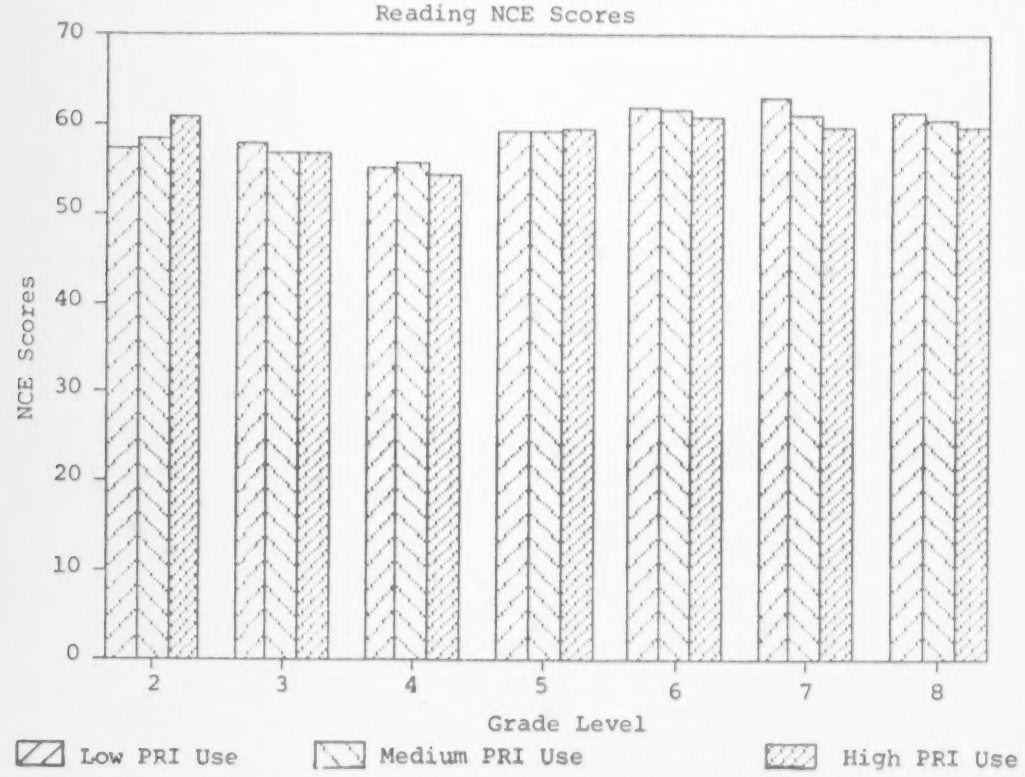


Table IV-28

Adjusted Mean CAT NCE Scores
for Mastery Ratio Groups

	Grade Level							
	2	3	4	5	6	7	8	All
Reading								
Zero Mastery	55.07	55.47	53.44	58.25	61.52	61.69	60.99	58.73
Below 50%	55.80	56.39	53.67	58.41	61.64	59.77	59.91	58.58
Above 50%	61.19	57.58	55.69	60.55	60.92	61.86	60.38	59.67
Full Mastery	61.76	57.92	57.41	60.99	61.72	61.68	60.08	60.19
Language Arts								
Zero Mastery		57.03	55.13	60.01	59.72	58.53	59.87	57.94
Below 50%		56.90	55.88	59.52	60.85	58.73	59.90	58.71
Above 50%		59.13	59.30	64.06	62.34	61.11	60.90	61.76
Full Mastery		57.75	58.86	62.24	63.42	58.58	61.67	60.71
Math								
Zero Mastery	61.27	58.71	56.32	59.46	61.40	61.57	61.63	59.83
Below 50%	60.50	58.95	54.22	60.91	61.04	60.76	61.92	59.65
Above 50%	63.52	60.64	58.24	64.38	61.99	61.40	63.42	62.15
Full Mastery	64.08	59.45	56.85	59.99	63.76	60.05	62.47	61.03

☐ Statistically significant differences ($p < .05$)

Table IV-29

Combined Adjusted Means for Students
Above and Below 50% Mastery Ratio

	Grade Level						
	2	3	4	5	6	7	8
Reading							
Below 50%	55.43	55.93	53.55	58.33	61.58	60.73	60.45
Above 50%	61.47	57.75	56.51	60.77	61.32	61.67	60.23
Language Arts							
Below 50%		56.97	55.46	59.81	60.28	58.63	59.88
Above 50%		58.44	59.08	63.15	62.88	59.84	61.28

The consistently high achievement of High Quality PRI Use students has positive implications for PRI implementation in Seattle Public Schools. Another highly positive finding is in the absence of statistical interactions between Quality PRI Use and the important student background characteristics studied throughout this evaluation. Trends toward high achievement among High Quality PRI Users are in equal evidence for all ethnic groups and socioeconomic levels in SPS. That is, High Quality PRI Use benefits low SES students as much as high SES students; or Black students as much as White students. Importantly, however, these statements are true based on the student membership in these Quality Use groups observed in the study. Although high Quality Use may benefit Black students as much as White students, the relative proportions of these students experiencing high and low quality PRI Use is an important issue in itself.

A general look at the ethnic and SES distributions among the four Quality PRI Use groups is given in Tables IV-30 and IV-31. Percentages of students in each of the groups are calculated based on the total number in that ethnic or SES subpopulation in the current data base. That is, the percentages across the four Quality Use groups within an ethnic or SES group will add to 100 percent. By looking at the pattern of these proportions within groups, one can see how students in a given group of interest (ethnic or SES) experienced varying levels of Quality PRI Use in 1983-84.

Across ethnic groups, only Asian students appear to evidence a different trend in Quality Use experience. They are the only ethnic group for whom 100% mastery is the most common experience, and also the only group for whom more than half of its students experience High Quality PRI Use (54%, as opposed to 45-46% in the other four ethnic groups).

Table IV-30

Ethnic Group Representation in Quality PRI Use Groups
Grades 2-8

		Asian	Black	Hispanic	Native American	White	Total
Zero Mastery							
	N	603	887	164	99	1825	3578
	%	21	25	25	25	25	24
At or Below 50%							
	N	694	998	189	119	2105	4105
	%	24	29	29	30	29	29
Above 50%							
	N	665	871	162	96	1790	3584
	%	24	25	25	24	24	24
Full Mastery							
	N	874	724	136	94	1612	3430
	%	31	21	21	21	22	23
Total							
	N	2836	3480	651	398	7332	
	%	19	24	4	3	50	

Table IV-31

SES Level Representation in Quality PRI Use Groups
Grades 2-8

		Low SES	High SES	Total
Zero Mastery				
	N	1414	2164	3578
	%	24	24	24
At or Below 50%				
	N	1605	2500	4105
	%	28	29	29
Above 50%				
	N	1556	2028	3584
	%	27	23	24
Full Mastery				
	N	1261	2169	3430
	%	22	24	23
Total				
	N	5836	8861	
	%	40	60	

High and low SES students share the same percentage of High and Low Quality Use, but differ in the breakdown within High Quality Use. Low SES has a majority of its High Quality Use students in the "greater than 50% mastery" group, while high SES students divide equally between this group and the Full Mastery.

As noted earlier, further dissection of these Quality PRI Use groups becomes an important concern when one tries to translate the positive findings in this factor into recommendations for implementation of the system. Further analyses of the composition of these groups, along a number of relevant student and PRI implementation characteristcs, are being conducted by the authors. To maintain the evaluative focus in the current study, results of these analyses are better left to a supplementary report.

Chapter V.

Evaluation Question 2: What are some Effective Practices in PRI Use at the Classroom Level?

The statistical analysis of the student data base has, in a sense, answered the "bottom line" evaluation issue in the study, i.e., has the use of the PRI/RS made SPS students better readers? Another question of the study is, given this evidence on the effectiveness/ineffectiveness of the system, how do teachers actually use the materials in their classrooms? Chapter III in the current report describes the implementation process for the PRI/RS as the system's authors intended it. Implementation evaluation research, however, has taught us at least two things in looking at the implementation of educational innovations. The first is that large, educational innovations are rarely implemented precisely as they were intended. Secondly, the fidelity of this implementation is a critical consideration in the overall evaluation. Simply put, if the program is not being implemented as intended, an evaluation of its effectiveness is not judging the merits of the program as it was developed, but as it is being used in the current context. Patterns of PRI/RS use in the current context, then, is an important consideration in this evaluation.

The current authors understand that the District has received much feedback from building level staff throughout the district on their concerns about the PRI. Some of this has been offered on a voluntary basis. More has been solicited recently through a series of planned interviews conducted by the District. In our study, we undertook a targeted, systematic inquiry into teacher and principal perceptions of the PRI/RS. Six schools were selected to participate in the interviews. All six were selected because they had

evidenced high use of the PRI/RS in its initial implementation year, 1983-84. Four were elementary schools, two were middle schools. Within these buildings, six teachers were chosen for interviews, again based on their recorded usage of the system in '83-'84. Among these six, we attempted to represent a range of PRI Use--three high and three low PRI users. Frequency of use at the school and teacher level were determined through our analysis of the student data base, and confirmed with the director of the TRACER office at the district. Individual teachers were identified through the identification number on the data base, and the TRACER coordinator was able to confirm that the selected schools were indeed high users of the system, and that the teachers selected represented the intended range of PRI use.

The protocol for contacting teachers and scheduling interviews began with the authors contacting our district liaison, who informed the elementary and middle school coordinators of the selected schools. He then telephoned the principals of the selected schools, described the purpose of the interviews, and indicated they would be receiving a call from the authors with information as to specific interviewees and possible dates. Principals' advice as to the the best time of day for each interview would be sought. The authors made these contacts, and arranged to spend three days in the district conducting the interviews. Letters were sent to each principal confirming dates and interviewees at each school.

Interviews were planned to take no more than one hour, and be conducted at a time and place of most convenience to the teachers. Most interviews lasted about 30 minutes. They were conducted before school, during teacher preparation periods, and after school in most schools to avert the need to employ substitute teachers. In two schools, however, substitutes had been arranged, so that interviews often lasted an hour and were scheduled in sequence.

There are two important interpretive guidelines to keep in mind when reviewing the results of the interviews. The first is that the findings of this portion of the evaluation study are not presented as representative of teacher perceptions throughout the district. As noted, schools were not selected for that purpose. They were selected to represent the most positive prospects for PRI/RS implementation. The authors hoped to learn more about effective uses of the system through this process. It seemed that the district had already been abundantly apprised of the problems with the system. Secondly, the comments of teachers and principals solicited through the interviews pertain to the first year of PRI/RS implementation, except when noted otherwise. Eliciting such retrospective information is sometimes difficult, but the presence of the PRI boycott in what would have been its second year of implementation actually helped in the process. That is, more recent experience with the system did not interfere with interviewees' recollection PRI Use in its first year in the district.

With these guidelines in mind, the authors present the findings of the interviews organized within four general areas which emerged from the 28 structured interviews:

1. Substantive Concerns about the PRI/RS
2. Classroom Uses
3. TRACER Reporting System
4. Relations with the District

Substantive Concerns

Teachers interviewed found the instructional activities in the Teacher Resource File to be filled with good ideas for teaching reading through the PRI objectives. Further, they found the PRI/RS objectives to be congruent with those they had been teaching in previous years in the SPS curriculum--a

finding corroborated by our analysis of the curricular match between the PRI/RS and the SPS reading curriculum (see Chapter VI of this report). As a resource, they found these materials extremely useful. This was particularly true for less experienced teachers, who were still developing their repertoire of instructional activities.

Many teachers noted an important aspect of the way in which the PRI/RS approaches the instruction of reading. They saw it as a discrete, skill by skill approach, as opposed to a more continuous, comprehensive approach. Many expressed the concern that this would sacrifice students' appreciation of the more literary aspects of reading, although it was very good for skills requiring drill and practice activities. This concern was particularly in evidence at the higher grade levels.

A few teachers raised the concern for an ethnic or cultural bias in the materials. Greater representation of Asian people in the reading passages or exercises was brought up by teachers in a school with a fairly large Asian population, for example. The system's publishers are very clear on their treatment of this issue of the representation of minorities, geographic regions, income levels, etc., in their materials. Their position is that these are represented in the proportions in which they actually exist in U.S. society. That is, if five percent of the nation's students are Asian, five percent of the instructional activities and test passages would contain references to Asian people. Although this general representation will not apply to all schools and grades in all regions, it is not generally viewed as a weakness in the PRI materials. The authors' analysis of the student data base presented in Chapter 4 does not support concerns for any ethnic or SES bias in the materials. The effectiveness/ineffectiveness of the PRI/RS is in equal evidence across all ethnic groups and SES levels in the SPS population.

Many teachers reported a lack of congruence between the instructional activities and the mastery tests for certain objectives. That is, they would teach the objective one way using the instructional materials, and it would be tested in a different way on the mastery test. These concerns were based on two very different and essentially independent characteristics--the format and difficulty level of the exercises. While both of these can put students at an unintended (and unfair) disadvantage in mastery testing, they are both avoidable from a teacher's perspective. Format differences can be remedied by developing activities which conform to the format of the mastery tests. This is not a form of "cheating" or teaching to the test. Format discrepancies are irrelevant distractors to the student, adding nuisance variation to test performance. To the extent they can be removed, test results will be more valid. Discrepancies in difficulty level were also encountered. Typically, mastery test items were more difficult than the instructional exercises. Again, teachers can deal with this, e.g., by increasing the difficulty of the practice exercises for some of the objectives.

Virtually all of the teachers expressed skepticism with the mastery decisions rendered by the PRI/RS, both in the IOI and in the mastery tests. They pointed out that they just didn't believe that complete mastery of an objective could be demonstrated on a four item (IOI) or six item (mastery test) test. They have observed that many students did not really know the material for some of the objectives the PRI/RS had indicated they had mastered. They typically agreed, however, with nonmastery decisions rendered by the system. Unlike all other substantive concerns with the PRI/RS, the authors concur with this objection on the part of teachers. From a measurement-theoretic perspective, there are limitations in the validity of a

mastery/nonmastery decision based on a small number of items. This is an ever-present point of contention in the areas of criterion-referenced and minimum competency testing. In multiple choice testing, there is always the element of chance operating in students' responses to the items. The probability of their reaching a criterion score on a set of items by guessing alone obviously increases with a small number of items. Implications of this for PRI/RS implementation likely involve suggestions for re-testing students for whom such conclusions are in doubt, in the opinion of teachers.

Summarily, substantive weaknesses of the PRI/RS are not a serious issue in SPS. Teachers we interviewed either felt positively about the system, or their concerns represented misunderstanding of the materials or their intended use. An exception is in the invalidity of some mastery/nonmastery decisions.

Classroom Uses

Virtually all teachers we interviewed indicated that using the PRI/RS represented extra work for them. In most cases, this was seen as a great deal of extra work. In other words, despite the good fit of the system with the SPS curriculum, it still took these teachers extra time to use the system well. Most examples were in the instructional materials. While they contained a wealth of good ideas, they did not go beyond the "idea" stage. Teachers had to develop the exercises themselves, rather than simply being able to take prepared worksheets out of the Teacher Resource Files, duplicate them, and use them in their classes. Again, this concern was more in evidence for the teachers who had not yet accumulated many of their own instructional materials. Other teachers who had such an accumulation found it less burdensome to simple adapt their materials, incorporating a perspective they had gained from the PRI/RS Teacher Resource File.

A related issue raised by a few teachers was in the time they had to spend in the testing and reporting functions of the PRI/RS and TRACER. They felt this was supplanting valuable instructional time. Further discussion in the interviews indicated this was more a comment on the TRACER system than on the administration of the mastery tests themselves. Particularly in the early grades, teachers spent an inordinate amount of time making sure the computer-readable forms were coded properly.

A frequent topic of extensive discussion in the interviews centered around teachers' classroom management strategies for effective use of the PRI/RS. In a sense, the PRI/RS is intended to be an individualized reading instructional program. This is, the IOI indicates which objectives each student has not yet mastered, the instructional activities help the student acquire the necessary knowledge, and the mastery tests assess his/her mastery of those skills. In theory, each student in the class could be at a different point in this process at any given time. Obviously, a teacher cannot be conducting 30 different reading programs for 30 students in his/her classroom. Some management and grouping strategies must be employed. In our interviews, we found some teachers resolving this issue very well. We found others needing assistance, and indicating that it was the one major drawback in using the system effectively.

There were distinct differences among grade levels in their classroom management strategies. At the early grades (K-2), when students were just beginning the reading process, many teachers found it useful to group their students into high, medium and low readers. These three groups would typically be working on different PRI objectives, and at different paces. Teachers were able to move effectively between these groups, occasionally

asking students in one group to help students in a slower moving group. This notion of peer tutoring was used increasingly frequently through the intermediate grades. In the middle schools, the instruction was often not as sequential as in the early grades. Few new reading skills are taught at this level. Instead, based on each student's status as per the PRI objectives, teachers would make mastery tests available during certain weeks of the school year. Students could decide when and how many of the tests to take (of those objectives they had not yet mastered), and turn them into the teacher. In a few cases, this testing occurred only at the end of the school year. With this flexibility in implementation, it's easier to understand some of the more outlying or unanticipated results of our analysis of the student data base as reported in Chapter IV (e.g., those students who had no tests submitted prior to Spring CAT testing, and the few students who had 60-70 tests submitted).

The availability of a support person in some of the schools we visited was seen as highly beneficial, primarily in terms of easing the burden and ensuring quality in the TRACER reporting system. Other schools often brought up the desirability of such support toward a less burdensome use of the system by teachers.

In relation to the previously discussed discrete-skill approach to reading represented by the PRI/RS, some teachers offered ways in which this was beneficial. Low achieving students, for example, profit from this type of approach. Students at any achievement level who evidence persistent problems in a more continuous form of reading instruction can profit from remediation at this skill level.

In the opinions of many of our interviewees, the utility of the system depended more on the characteristics of the teachers than the students using the system. Teachers with effective implementation strategies for the PRI/RS

acknowledged that it was more work, at least initially, to use the system. Those that described themselves as appreciating an organized, accountable system for keeping track of students' progress felt it made a contribution to their teaching. They suggested that teachers who were not used to keeping track, or organizing their instruction, opposed the system. One teacher put it rather bluntly: "Good teachers like it. Bad teachers don't."

Summarily, effective use of the PRI/RS involves a significant amount of extra work and requires effective classroom management strategies. Our sample interviews suggest there is a wide variety of effective strategies in use in the district.

TRACER

One of the more difficult challenges we faced in the interviews was to separate teachers' perceptions and opinions about the PRI instructional and assessment system from those of the computerized reporting system known as TRACER. In reviewing previous reports of PRI use in SPS, and in discussions with district staff, the TRACER seemed to be more the focus of school and teacher complaint than was the PRI system itself.

We found there to be no shortage of such complaints in our interviews. A list of specific problems offered by the 28 teachers we interviewed would go on for pages, and it was not our purpose to troubleshoot the reporting system. We understand that the district has gone to great lengths to revise and improve the system since its inception. We can, however, corroborate the general impressions of the system in the schools we visited, and perhaps offer our perspective from much experience with the implementation of technological innovations on an audience without much technical background.

During the first year of PRI/RS implementation, TRACER was viewed as excessively burdensome on teachers, filled with errors in its processing, insufficient in ongoing technical assistance and highly redundant in its output. From a teacher's point of view, it took an undue amount of time to fill out and check for accuracy the required forms in the input phase of the system. Again, those schools that had a support person available to take over this responsibility were not as vocal about this complaint. This burden, combined with persistent errors in the resultant output, easily diminished their interest in complying with the TRACER system. Further, many teachers reported serious lack of responsiveness to their questions or concerns for errors on the part of the central TRACER staff. Calls were frequently not returned, or, when they were, the response was always that the teachers were at fault.

There were important differences in perceptions of TRACER in the first and second year of its implementation. Interviewees noted some improvement in the coded data entry process, and in the more careful selection of relevant output for teachers. Concerns for persistent errors in the system, and more ongoing technical assistance remained however.

In a very real sense, the effective implementation of the TRACER system is the subject of an evaluative effort all its own. The district has taken steps to ease the coding requirements for teachers. More are needed. The district has offered technical assistance to schools and teachers voicing concerns about the system. More are needed--from a client-centered perspective. The district has eliminated some of the redundant reports, and added others based on interest expressed by teachers. More of this specific tailoring of information is needed. All of these suggestions, and consequent efforts by

the district, must be taken with the inevitable imperfections of such a highly computerized system in mind. That is, such systems never operate flawlessly or in a way that meets everyone's needs. Often, however, with enough training and ongoing technical assistance, users become invested in the utility of the system, and are not so focussed on its limitations. We understand that the district has hired an external consultant with expertise in systems analysis. We support this addition to the overall evaluation/revision/renewal cycle of TRACER implementation.

Summarily, in its first year the TRACER was viewed as a burdensome, error-filled system, whose output was not worth the effort it took to use it.

Relations with the District

Although it has been more than two years since the PRI/RS was introduced in the district, there are still strong feelings of opposition to the way in which it was introduced. Nearly all of the teachers we interviewed felt they had no representation in the adoption decision. The absence of a pilot phase of implementation in a few schools (as has since been done with the Diagnostic Math Inventory) was seen as a serious omission in the process. With this image of a burdensome system being thrust upon them with no input and little warning, it is easy to view the PRI/RS as an expensive and unnecessary innovation. Many teachers we interviewed felt exactly that way. Again, they saw some merit in the PRI/RS, as has been reported, but did not agree with the way in which the initial decision was made.

As discussed in the previous section on TRACER, any innovation requires initial training and ongoing technical assistance. The sequence of training offered in conjunction with the adoption and early implementation of the

PRI/RS is reported in the Introduction section (Chapter I) of this report. Interviewees indicated that it was not adequate. They criticized the quality of the training, particularly that conducted by the PRI/RS authors. Specifically, they felt these presentations were condescending and out of touch with actual classroom instruction. They suggested that attendance was low, not reaching enough of the ultimate users in the district. Ongoing technical assistance has also been insufficient. The need for user groups sharing ideas was brought up by several teachers.

Summarily, there is a significant lack of motivation and decay in teacher morale due to perceptions about the initial decision-making processes and insensitivity to ongoing needs for support.

Chapter VI.

Evaluation Question 3: Does the Curriculum of the PRI/RS Match that of SPS and CAT?

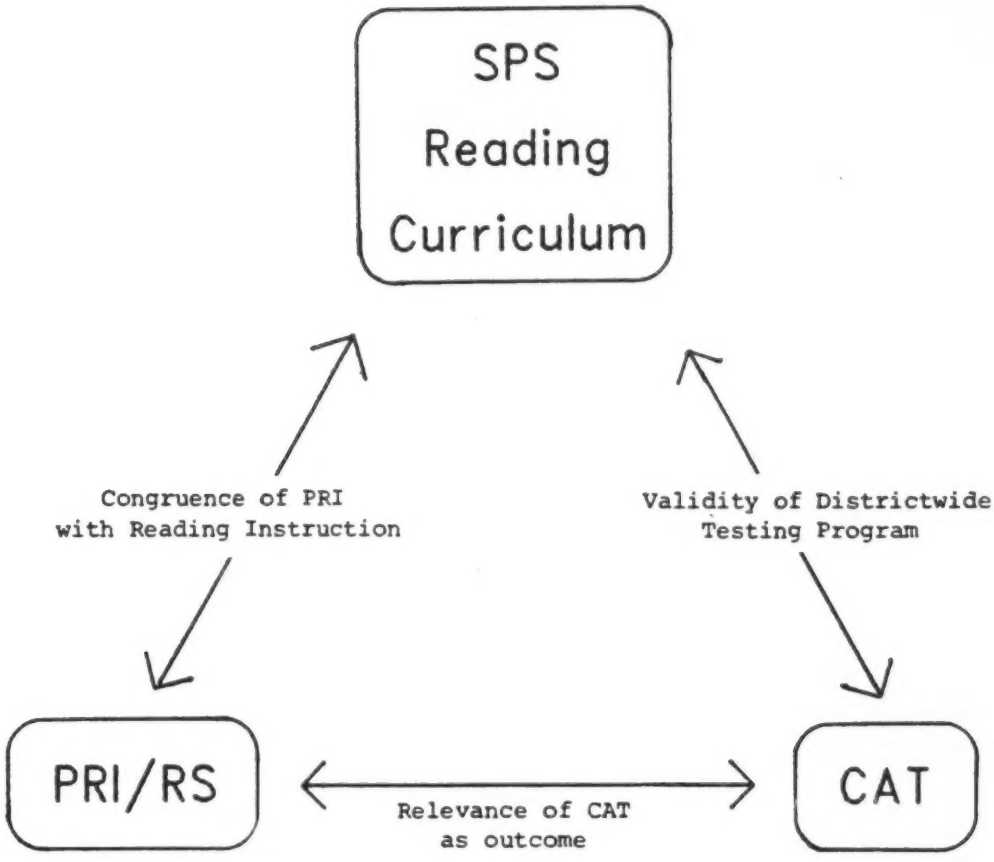
The value of the PRI/RS as a diagnostic assessment tool for Seattle Public Schools is naturally limited by the extent to which the PRI measures the reading skills taught in classrooms in the district. The instructional component of the system, in a sense, guarantees this. Part of the value of the system, as well as its potential for success, rests on the match between its curriculum and that already in place in SPS. If these are reasonably congruent, the PRI/RS offers simply a new approach to teaching the skills already in place in SPS. If they are not, the system represents more than a new instructional/assessment approach. It represents a new curriculum.

The third major activity of the current evaluation was to conduct a thorough analysis of the content of these two curricular system. In addition, the content of the districtwide test, the California Achievement Test (CAT), is an important consideration in this evaluation study. The CAT was selected as the criterion measure for the evaluation of effectiveness of the PRI/RS in the distict. It too must possess a reasonable match to the PRI and the SPS curriculum if it is to be sensitive to relevant differences in student achievement.

The interplay of these three curricular components is displayed in Figure VI-1. The SPS Reading Curriculum appears at the top of the diagram because it is the basic element from which assessment devices are selected. The curricular validities of both the PRI and CAT are pre-eminent considerations in judging the utility of those assessment devices in SPS. The content match between the PRI and the CAT determines the relevance of the CAT as an outcome measure in the current study.

Figure VI-1

Relations Between Three Curricular Components
in Evaluating the Effectiveness of the PRI/RS



The arrows connecting these three components in Figure VI-1 describe these relationships and the specific issues with which they deal. Of most relevance to the current evaluation is the curricular match between the PRI and the SPS Reading curriculum. The outcome of this analysis will determine the extent to which the PRI was an innovation primarily of instructional and assessment methods, or one of curricular implications. The contention of the district has always been the former. In addition, the match between the PRI and the CAT will be discussed as it relates to the current evaluation. The match between the SPS curriculum and the CAT will not be assessed here, however. As indicated in the diagram in Figure VI-1, it is an issue which the district considered heavily in its adoption of the CAT for its districtwide testing purposes many years ago. While important for the validity of that testing purpose in the district, this relationship does not play an important role in the current evaluation.

In assessing the match between the PRI/RS and the Seattle curriculum, two important questions arise. One concerns the extent to which the new system (PRI/RS) covers the existing curriculum (SPS). It can be addressed by reviewing all objectives in the SPS curriculum at each grade, and determining how many of these are also represented in the PRI/RS. We will refer to this as the comprehensiveness of coverage issue. A second question concerns the extent to which the PRI covers only skills which are contained in the existing curriculum. This is a relevance of coverage issue. The two are similar, but not identical. For example, the PRI may be very relevant to the SPS curriculum, but not comprehensive. That is, all PRI objectives at a given grade may have their counterparts in the existing curriculum; but there still may be several SPS objectives at that grade which are not covered by the PRI. Thus, these two issues must be addressed separately.

Using curriculum maps of the PRI/RS furnished by CTB/McGraw Hill and the SPS Reading Curriculum guides (K-6) furnished by the district, the authors analyzed the match of these two curricula at the instructional objective level. A tally of matched objectives was done at all combinations of grade levels (SPS) and test levels (PRI). Results of this analysis from both perspectives discussed above are contained in Tables VI-1 and VI-2.

In Table IV-1, a summary of the number of matched objectives is presented in relation to the comprehensiveness issue. The number (and proportion) of SPS objectives tested and not tested by the PRI are shown in the Table for publisher-recommended grade and test level combinations. For example, PRI Level A includes 26 of the 40 reading objectives in the SPS Kindergarten curriculum--65 percent coverage. At grade 1, where the PRI publishers indicate either Level A or Level B can be used, the comprehensiveness of coverage is very different. Level A covers only 28 percent (13 of 46) of the SPS reading objectives, while Level B covers 67 percent (31 of 46). On this basis it is apparent that Level B is the far better match to Seattle's curriculum at grade 1. As noted earlier, a high percentage of SPS students in grade 1 took Level A of the IOI (88%, see Table IV-3). This is not inappropriate, however, as the IOI functions as a pretest at the beginning of the school year. If students did not progress to Level B during the course of the school year, one could be concerned about the congruence of PRI instruction with that based on the district curriculum. Since a large proportion of students at grade 2 took Level B of the IOI (72%), however, this does not appear to be an issue.

The discrepancy in comprehensiveness of coverage at grade 1 is the only example of possible mismatch between a PRI level and the existing district curriculum. All other proportions are in the 60-70 percent range. It is not

Table VI-1

Number and Percent of SPS Reading Objectives
Tested and Not Tested on the PRI/RS

PRI Level...		A		B		C		D		
Grade...		K	1	1	2	2	3	4	5	6
SPS Objectives Tested	N	26	13	31	40	36	45	44	50	49
	%	65	28	67	71	64	73	69	75	71
SPS Objectives Not Tested	N	14	33	15	16	20	17	20	17	20
	%	35	72	33	29	36	27	31	25	29
Total SPS Objectives		40	46	46	56	56	62	64	67	69

Table VI-2

Number and Percent of PRI/RS Objectives
Taught and Not Taught in the SPS Reading Curriculum

PRI Level...		A		B		C		D		
Grade...		K	1	1	2	2	3	4	5	6
PRI Objectives Taught	N	11	10	23	29	26	35	31	31	31
	%	85	76	59	74	61	81	74	74	74
PRI Objectives Not Taught	N	2	3	16	10	17	8	11	11	11
	%	15	23	41	26	39	19	26	26	26
Total PRI Objectives		13	13	39	39	43	43	42	42	42

surprising that such a discrepancy might exist in the first grade. As has been noted in other analyses in the current evaluation, this is the grade level at which pre-reading and reading instruction often both occur.

Results of the content analysis of relevance of the PRI to the SPS Reading curriculum is contained in Table VI-2. At first grade, for example, 10 of the 13 instructional objectives on the PRI have counterparts in the existing SPS curriculum. In other words, 76 percent of the PRI Level A objectives are relevant to SPS instruction. As a total instructional package, Level B is less relevant at grade 1--only 59 percent of its objectives are also found in the SPS curriculum. How does this relate to the previous discussion of comprehensiveness of the PRI at grade 1, where Level B appeared to be superior to Level A? There are three times as many objectives on PRI Level B as on Level A. While its proportional relevance is lower than that of Level A at grade 1, it has a greater number of objectives in common with the SPS curriculum (23, as opposed to 10 for Level A). Still, if Level B were used exclusively at grade 1, there would be a higher incidence of irrelevant material covered than would be the case with Level A. This illustrates the meaningfully different perspectives on the content match issue taken by the comprehensiveness and relevance questions.

Summarizing the remaining values in Table VI-2, and viewing the "best" PRI levels at each grade, the relevance of the PRI ranges from 74 to 85 percent. These figures are slightly higher than those in the comprehensiveness analysis of Table VI-1. This is likely due to the wider breadth of content represented in the SPS curriculum than in the PRI/RS at all grade levels. Proponents of the PRI/RS would not view this as a weakness of that system. Going back to its portrayal as a discrete-skill approach to reading, its purpose is necessarily more limited than a curriculum that includes emphasis on application and appreciation of reading, particularly in the higher grades.

The match between the PRI and the CAT is important in affirming the relevance of the CAT as an indicator of PRI effectiveness in this evaluation. An analysis of the match of objectives assessed by both tests based on literature obtained from CTB McGraw-Hill (the publishers of both the CAT and the PRI) yielded 100 percent match from both comprehensiveness and relevance perspectives. Given the different purposes of the two assessment devices, and the consequent differences in their curriculum specifications, this "perfect" match is not surprising. The California Achievement Test is a survey assessment tool, designed to determine students' knowledge of a broad universe of reading skills and concepts. Its content is specified at the level of category objective--a more general level than the instructional objective unit of the PRI. A CAT level may consist of 5 or 6 category objectives, while the corresponding PRI level would contain 30 or 40 instructional objectives. If the match is assessed at the category objective level of both instruments, this level of generality easily yields perfect match.

Summarily, the analysis of the content of the PRI and the SPS curriculum suggests adequate match, such that PRI instruction can be viewed as complementary to that needed for the existing district curriculum.

Chapter VII.

Summary

The evaluation of the PRI/RS in Seattle Public Schools consisted of three activities---a statistical analysis of a large SPS student data base from the 1983-84 school year, a series of interviews with 28 teachers and principals in SPS, a content analysis and comparison of SPS Reading curriculum, PRI/RS instructional objectives and CAT category objectives. The statistical analysis of the 31,000 student data base was, by far, the major thrust of the evaluative effort reported here. Personal interviews added a narrative flavor to the interpretation of the results of the statistical analysis.

The findings are summarized in three major categories:

Levels of Use - based on statistical analysis of PRI Use as recorded in TRACER.

Impact - based on statistical analysis of demographic, PRI Use, and CAT achievement data in SPS data base.

Teacher Perceptions - based on teacher/principal interviews.

Levels of Use of the PRI/RS

A range of IOI levels were administered at each grade. Results suggest they were at an appropriate level of difficulty.

A substantial number of students took no mastery tests at all prior to spring CAT testing, particularly at the higher grade levels.

Through grade six, SPS students average one mastery test submitted per month. Middle schools' average is less than half of this.

Mastery testing occurred at a faster rate after spring districtwide testing than before at grades 7 and 8.

The distribution of mastery tests submitted is highly positively skewed, indicating a small portion of students for whom an extremely large number of tests were submitted (e.g., 60-70).

Different standards for high, medium and low levels of PRI Use existed between elementary schools (grades 1-6 in the study data base) and middle schools (grades 7-8).

Impact of the PRI/RS on Student Achievement

High users of the PRI/RS (more than one mastery test submitted per month before spring CAT testing) showed higher CAT Reading scores than medium or low users of equal prior reading achievement at the second grade.

In seventh grade, low PRI users evidenced higher CAT Reading scores than medium or high users of equal prior achievement.

At grades 3, 4, 5, 6, and 8 there were no reading achievement differences among high, medium and low PRI users.

In Language Arts, high PRI users demonstrated higher achievement than medium or low users of equal prior achievement at grades 3 and 5.

Using an index of quality rather than quantity of PRI Use, achievement comparisons evidence a more consistent pattern across the grades.

At all grades, students mastering more than half of the objectives on which they were tested showed higher CAT Reading or Language Arts scores than those mastering less than half. Previous differences in achievement between these two groups of students were statistically removed before making these comparisons.

Achievement differences among high, medium and low quality PRI users are evidenced in Reading in the lower grades, and Language Arts in the intermediate grades.

PRI Use measures showed moderate to low correlational relationships with 1984 CAT Reading or Language Arts scores. From a prediction standpoint, they added almost nothing to what could be predicted from the 1983 CAT scores.

All achievement differences noted among high, medium and low quality PRI Use groups are consistent across ethnic groups and socioeconomic status.

Teacher Perceptions of PRI Use and Effectiveness

PRI instructional materials contain a wealth of good ideas. They require a great deal of extra work on the part of teachers to prepare and put into practice, however.

Teachers feel that mastery criteria on the IOI and mastery tests err on the liberal side. That is, they observe that many objectives which have been mastered via testing still need remediation for some of these students.

Concerns for ethnic or cultural bias in PRI materials were not borne out in the analysis of student test scores. Students of all ethnic groups and socioeconomic levels profit equally from PRI Use in terms of reading achievement.

There are a few instances (i.e., objectives) of mismatch between PRI instructional activities and mastery tests.

The traditionally wide chasm between readers and non-readers at the first or second grade levels makes PRI implementation particularly difficult at these early grades.

There is a great deal of variation in the way teachers use the PRI/RS. Primary grade level teachers use the materials very differently from middle school teachers.

There is still opposition among building level staff to the way in which the PRI/RS was initially adopted. They view the initial training (CTB McGraw-Hill) as very poor and insensitive to their needs; and the ongoing support from the District as inadequate.

Teachers need ideas for effective classroom management strategies for PRI Use. Interviews in the current study indicate there are several already in use in Seattle Public Schools. Communication and sharing among teachers is needed.

The TRACER System is viewed as cumbersome and error-filled. Often, the PRI/RS is not distinguished from TRACER in teacher opinions about the system.

The content match between the SPS Reading curriculum and the PRI/RS instructional objectives is very good, exceeding that demonstrated in other studies of test/curriculum match across the nation.

The PRI/RS represents a discrete skill approach to teaching reading. This is in contrast to a more continuous and comprehensive approach.

The authors recognize that the District has undertaken several evaluation-related efforts of the PRI/RS since its adoption prior to the 1983-84 school year. Initial reactions to the results reported here have spawned a series of additional analyses to be conducted on the current data base. Such secondary analyses will be topics of subsequent reports by the authors. We view the findings of the study reported here as firm evidence of the effects of the PRI/RS in SPS and hope their presentation assists in policy decisions to come.

Appendix A

Study Design

An Evaluation of the Use of the Prescriptive Reading Inventory/Reading System in Seattle Public Schools

STUDY DESIGN

Prepared by

Dr. Gary D. Estes
Dr. Roy M. Gabriel
Assessment and Evaluation Program
Northwest Regional Educational Laboratory
300 S.W. Sixth Avenue
Portland, Oregon 97204

December, 1984

Amended as indicated,
January, 1985 -RG

CONTENTS

Major Section	Page
Overview of Study Design	1
Activity 1: Test Data Analysis	3
Activity 2: Test/Curriculum Match	17
Activity 3: PRI Implementation Survey	19
Activity 4: SPS Staff Interviews	21
Summary	23

Tables

Variables Used in the Analysis of 1983-84 Data	8
Projected Sample Sizes for Longitudinal Cohort Analysis	11
Contents of a Student Data Record	14
Critical Events	24

Figures

Illustrative Two Factor ANOVA Design	7
Possible Longitudinal Study Design	10

The fundamental question addressed in this study may be phrased quite simply as "Are students in Seattle Public Schools (SPS) better readers as a result of the use of the PRI/RS materials in the district?" Following this general question, there are several subsidiary questions relating to differential effects on reading achievement according to a variety of student and PRI usage characteristics. Still another set of issues of import to this study is in the specific implementation of the PRI/RS in Seattle Public Schools. The design of this study will focus on these two general areas: the implementation of the PRI/RS and its effects on the reading achievement of students in Seattle Public Schools.

Overview of the Study Design

To inform the design of the study, the authors have met with SPS staff on several occasions and have contacted CTB/McGraw-Hill test publishers and a number of public school and district staff across the country who use the PRI via telephone interviews. These interviews have included staff from the CTB main office in Monterey, California and a regional office in Atlanta, Georgia. District and building level staff have been contacted in the following locations: Akron, Ohio; Bakersfield, California; Memphis, Tennessee; Oklahoma City, Oklahoma; and Wilkesborough, North Carolina. These districts represent as much as eight years use of the PRI/RS and their range of experiences have contributed valuable information to the design of the current study for SPS. Meetings with SPS staff have also been extremely informative in representing the history of PRI adoption and implementation, orienting the authors to district operations of relevance to the current study, and clarifying the important questions of the study. Taken together,

these preliminary activities have helped formulate a design which is targeted to SPS primary converns and is sensitive to factors which, in other district's experiences, may influence the basic study questions in years to come.

There are essentially ~~four~~ three activities to be done in the study:

1. Test Data Analysis - a series of analyses of test results obtained from the PRI and the California Achievement Test (CAT) to assess the relationship between PRI use and reading achievement. These will include subsidiary analyses by student characteristics.

2. Test/Curriculum Match - a review and comparison of the instructional objectives in the PRI, CAT, and SPS curriculum to determine the goodness of fit of these materials at a content level.

3. ~~PRI Implementation Survey - a questionnaire sent to a sample of building level staff throughout the district with questions regarding the implementation of the PRI/RS.~~

4. SPS Staff Interviews - a series of interviews with building level staff in the district designed to probe more specifically patterns of attitudes toward PRI use in the district.

Specific plans and procedures for these four major study activities are detailed in the remainder of this Study Design proposal. Hopefully, they will not be conducted as four separate and independent activities. Rather, their findings will supplement and, in some cases, guide each other to produce a comprehensive evaluation of PRI/RS in Seattle Public Schools. The extent to which this ideal can be attained may be largely a function of the management of the study within timelines of ongoing operations in the district. Therefore, this study design includes specifications of timelines, data sources and needs for SPS staff assistance in each of the activities.

Replicate these analyses on '84-'85 school year
-RG

Activity 1: Test Data Analysis

The primary activity in this study is a thorough statistical analysis designed to assess the effects of PRI/RS in the district on reading achievement of its students. This analysis will employ two basic designs:

a. Achievement Status Design - Analysis will focus on the 1983-84 school year only. Data will include student characteristics, PRI usage and results, and CAT reading achievement results.

b. Longitudinal Achievement Design - Analysis will highlight previous achievement trends of student cohorts identified in Status Design above.

The first analysis conducted will concentrate on data collected during the 1983-84 school year, the first year of PRI/RS implementation in the district. Individual student data will be needed from three district data bases at SPS: the districtwide CAT test history file (TH), the Student Placement Activity Master file (SPAM), and the TRACER file. Specific data needed for the analysis in this activity are detailed at the conclusion of this section.

Initially, data from the TRACER and SPAM data files will be analyzed to determine levels of use of PRI by grade level throughout the district. A frequency distribution of the number of objectives mastered during IOI diagnostic testing (i.e., fall "pretest") will be constructed at each grade level. This "pretest" analysis is useful in detecting any possible floor or ceiling effects in the PRI level administered. If levels of PRI use are to be determined largely by the number of objectives tested or mastered during the year, these initial scores must be sufficiently low to allow such growth to be demonstrated. The detection of floor or ceiling effects at the diagnostic stage of testing not only affects subsequent analysis of PRI usage and its effects on reading achievement, it may suggest improper use of the materials themselves, i.e., administering an inappropriate level of the test to a particular student. Both of these are of concern in this study.

Next, the number of objectives tested and mastered during the remainder of the school year will be analyzed in the same fashion. These analyses, along with those of the IOI diagnostic tests, will empirically determine the definitions of "high," "medium" and "low" PRI use for subsequent analysis. In preliminary meetings with SPS staff, cutoff values for these categories have been suggested (e.g., 10 or more objectives mastered during the year represents "high" PRI/RS use, 5 to 9 is "medium" use, 4 or less is "low" use), but the authors suggest this is best addressed empirically. The frequency distributions constructed here will also determine whether the degree of PRI use is roughly equivalent across grade levels. It will be advantageous in the interpretation of subsequent analyses to have the same definitions of "high," "medium" and "low" PRI use across grade levels. This will be an important goal for the authors in designating cutoff values for these levels of use categories.

Correlational and scattergram analysis within grade level will be used to assess the relationship between the number of objectives mastered at pretest and the additional number mastered during the remainder of the school year. Often, this type of analysis yields a highly negative correlation--that is, the more a student knows at the beginning of the year, the less he or she has to learn during the rest of the year. Other outcomes are possible, however, particularly if there are no floor or ceiling effect problems in the instruments. The results of this correlational analysis will be informative both in interpreting the PRI "gains" during the school year, and in viewing its effects on reading achievement.

Following the initial analysis of PRI data as contained in the TRACER file, its relationship to student reading achievement will be determined using the CAT results supplied from the SPS Test History file. The basic question of the relationship between PRI use and reading achievement can be addressed

Add math & language arts
CAT scores to this analysis -RG

statistically in two general, and related, ways. They will be described briefly here, noting that these two approaches have advantages and disadvantages which vary with the number of additional variables to be included in the analysis (e.g., student gender, ethnic origin).

The first approach employs a statistical Analysis of Variance (ANOVA) model, examining CAT reading scores broken by categorical factors such as ethnic group and levels of PRI Use. Figure 1 displays a sample factorial design for such an analysis. The ANOVA approach consists of comparisons of means of CAT reading scores according to the categorical breakdown in the factorial design. The simple design shown in Figure 1 addresses two of the important questions in this study:

1. Is reading achievement significantly different for students in the high, medium and low PRI use groups? (A Usage main effect.)

2. Is the pattern of reading achievement for different levels of PRI use the same for all ethnic groups? (A Usage by Ethnic interactive effect.)

It should be noted that this analysis can also address the question of CAT reading achievement differences among the five ethnic groups, i.e., an Ethnic main effect. These achievement differences are already well documented in reports produced by the SPS testing office describing districtwide test results (e.g., Report No. 84-5, Jan. 1984), however, and are not the focus of the current study. Differences in PRI use or in its impact on reading achievement are central issues in this study, and these are addressed through prior TRACER analyses or in the "Usage by Ethnic" interaction specified above.

The advantages of the ANOVA approach are in the relatively clear and straightforward test of hypotheses of interest. Comparisons of means can be presented quite simply in tables or graphs. In some applications, however, this clarity is gained at the expense of more specific or accurate representation of variables. Such may be the case for the PRI Use factor in

this study. Categorizing this variable, which is actually measured by the number of objectives tested or mastered during the course of the year, into three general categories may sacrifice information and obscure any differences in reading achievement which may occur within one of these categories. Hopefully, the preliminary analysis of TRACER data previously discussed which will determine the cutoff values for these categories will minimize this possibility. In addition to this potential shortcoming, the ANOVA approach can become unwieldy if a large number of factors are of interest and are to be examined in combination in a design such as that shown in Figure 1. Imagine, for example, expanding the illustrative design to include student gender, SES, and grade level. The cells of the design become so complex (e.g., third grade, male Asian students who qualify for free lunch and exhibit a high level of PRI use) and comparisons so minute that the fundamental issues in the analysis can become lost.

When there are a large number of variables of interest, such as in this study, their effects on reading achievement may also be assessed using another statistical approach, Multiple Linear Regression (MLR). This approach employs correlational analysis in which an outcome variable (reading achievement) is related to a set of predictor variables (e.g., PRI use, ethnic group, SES, initial achievement level, etc.). The product of this analysis will be essentially twofold: a single correlation coefficient which indicates the strength of the overall relationship between CAT reading and the entire set of predictor variables; and individual weights for each variable indicating their relative influence in the overall relationship with reading scores. While MLR results are not as immediately interpretable as those of ANOVA, they do address the general study question in a precise, statistical way, i.e., "Does use of the PRI have a positive effect on reading achievement, and is the effect the same for different types of students?"

Figure 1

Illustrative Two Factor ANOVA Design

Use of PRI/RS

Ethnic Origin:	High	Medium	Low
Hispanic	(CAT Reading Scores)		
Native American			
Asian			
Black			
White			

While the ANOVA and MLR models appear quite different on the surface, they are highly related statistical techniques, usually yielding the same conclusions. Used in tandem, their relative advantages can be brought to bear on a comprehensive analysis of the data at hand. This will be the strategy employed by the authors in the analysis of the 1983-84 data. Initially, MLR will be used to analyze the effects of the host of variables of interest in this evaluation on CAT reading scores. Interpretation of this analysis will identify those variables which affect reading and those which don't. Once identified, these factors can be treated separately or in smaller sets via simpler analyses such as the ANOVA model or graphs. A list of the variables involved in the 1983-84 analysis is contained in Table 1.

The variables and analyses described above represent a complex, but thorough analysis of the basic study question for a single school year. It is the intent of the authors that this approach could be replicated by SPS staff or other contractors in future years in order to maintain an ongoing annual evaluation of PRI use in the district.

Table 1

Variables Used in the Analysis of 1983-84 Data

Outcome	Predictor
1984 CAT Total Reading	Level of PRI Use
1984 CAT Reading Vocabulary	Ethnic Origin
1984 CAT Reading Comprehension	Gender
Add: 1984 CAT Total Math	Socioeconomic Status (free/ reduced lunch)
Math Computation	Grade Level
Math Concepts & Applications	Bilingual Fluency
CAT Total Language Arts	Special Education Participation
LA Expression	Chapter 1 Participation
LA Mechanics	

-RG

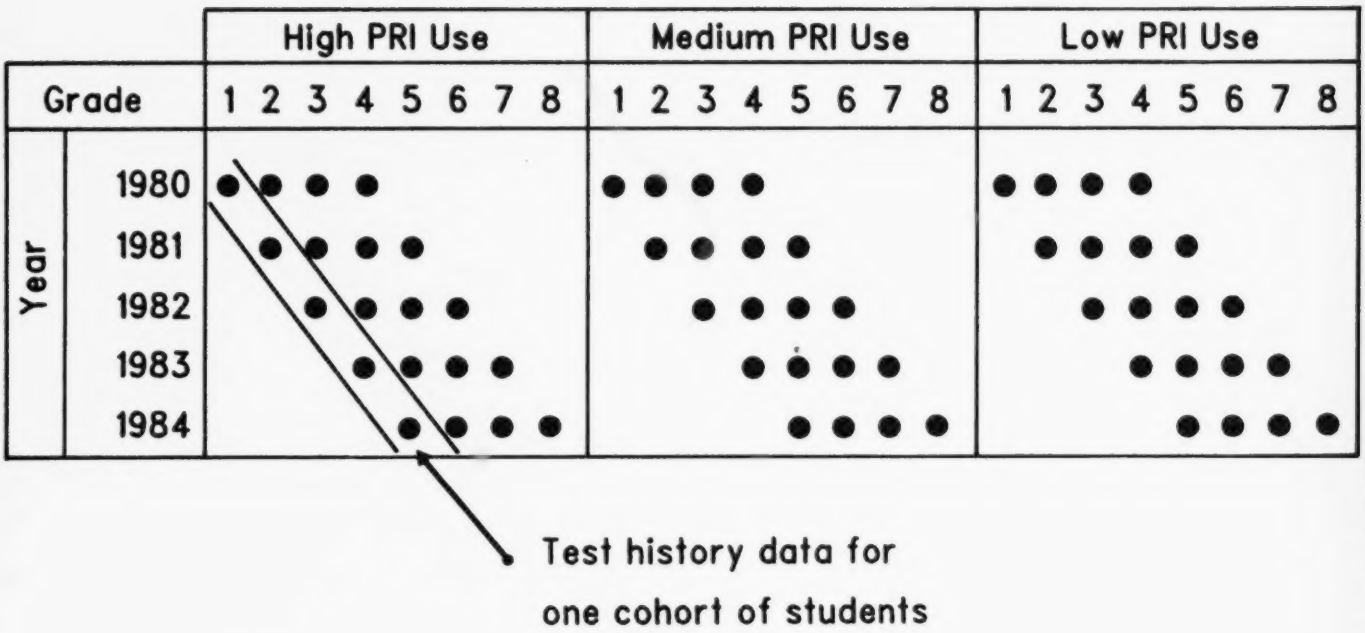
The second general thrust of test data analysis described at the beginning of this section involves tracing the history of reading achievement of selected cohorts of students identified in the 1983-84 analysis. This longitudinal analysis addresses an important issue in the basic study question, but is vulnerable to a host of threats to validity common to all longitudinal efforts. These will be mentioned during a basic description of the longitudinal analysis designed for the current study.

From the 1983-84 analysis, three cohorts of students at each grade level will be identified--those evidencing high, medium, and low PRI use. The achievement of these cohorts at the Spring, 1984 testing (after a year of PRI use) may be compared to a baseline of achievement beginning as early as 1980. This is displayed graphically in Figure 2.

The number of years in the pre-PRI baseline is an important decision. Districtwide spring CAT test summaries are available since 1980 (SPS Report No. 84-5; Jan. 1984), providing important district norms for achievement trends against which those of the PRI use cohorts can be compared. The farther back in time one goes, however, the smaller will be the sample size in each cohort. This is simply a realistic consequence of student mobility and absences from testing in typical district operations. The authors have conducted some preliminary analysis of projected sample sizes available at each grade level for longitudinal analyses of varying length, i.e., 2 to 5 years. Based on information in the districtwide testing report, roughly 75 percent to 85 percent of students have test data available in two consecutive years. Using total sample sizes (excluding bilingual students) for the Spring 1983 testing and the percent of matched cases from the 82-83 analysis, as estimates, projected sample sizes for two through five year longitudinal analyses are given in Table 2. Given the importance of representing the original PRI usage cohorts adequately, and the need to

Figure 2

Possible Longitudinal Study Design



further subdivide those groups by relevant characteristics (e.g., Ethnic Group, SES), the authors recommend a two year longitudinal analysis. Using 1983 data as the pre-PRI achievement data will likely retain 75 percent of the original sample, while going to earlier years reduces this to 53%, 37%, and 22%, respectively. Although these percentages are only projections, they are likely optimistic ones, and therefore the authors highly recommend the use of 1983 CAT results to represent reading achievement prior to PRI implementation.

Table 2

Projected Sample Sizes for
Longitudinal Cohort Analysis

	1	2	3	4	5	6	7	8	
1980	1064 (38)	900 (43)	1363 (47)	1250 (50)					N5= 4,577 (22%)
1981	1242 (46)	1400 (50)	1710 (57)	1711 (59)	1525 (61)				N4= 7,588 (37%)
1982	1450 (58)	1647 (61)	1848 (66)	2130 (71)	2088 (72)	1750 (70)			N3=10,913 (53%)
1983	1900 (76)	1900 (76)	2160 (80)	2296 (82)	2610 (87)	2407 (83)	2110 (84)		N2=15,383 (75%)
1984	2600	2500	2500	2700	2800	3000	2900	2500	N1=20,500

Using one year of test data as the pre-PRI achievement status indicator allows a number of options for data analysis. In the general ANOVA model previously described, 1983 CAT data could be used as a covariate in analyzing the effects of the design factors (PRI use levels and ethnic origin). A repeated measures ANOVA could also be employed, in which Year of Test Data (1983 or 1984) is a design factor. Still another option would be to classify students into high, medium and low Previous Achievement Level groups, and use this as another design factor. The selection of the method of choice among these is highly dependent on data characteristics such as the magnitude of the correlation between 19893 and 1984 CAT scores, its equivalence across grade levels and other design factors, and the extent to which "high, medium and low" classifications are a natural division of the 1983 test data.

As in the analysis of the Achievement Status Design, MLR is another data analytic approach here. It would be a simple extension of that presented in Table 1. CAT Reading scores from Spring 1983 would be added as predictor vairables to the analysis.

Timelines/Data Needs

The work outlined in this section requires data from three SPS data bases: CAT Test History, SPAM, and TRACER. Preliminary meetings with SPS data processing and testing office staff and a review of these file structures by the authors have indicated that the data referenced in this section are available in any or all of these files. The requisite data from these three data bases must be extracted and merged into a single data file. This will allow the analysis prescribed in this section to be conducted in a timely manner, ensuring that the same student population is represented in all analyses. Table 2 contains a list of variables needed for each student. There are four general clusters of variables: a unique student ID number

common to all SPS data bases, demographic characteristics taken primarily from the SPAM data base, PRI objective mastery data taken from the TRACER data base, and CAT reading scores from the testing office Test History data base. It is hoped that much additional clarification of variable definitions, tape specifications, etc., can be completed as part of the discussion of the study design.

Table 3

Contents of a Student Data Record

	Variable Cluster	Variable Name	Source	Field Length/ Possible Value
1.	Student Identification	Student ID	TH, SPAM, TRACER	7 digit
2.	Student Demographic Characteristics	1983-84 School ID	TH, SPAM, TRACER	3 digit
		1983-84 Teacher ID	TRACER	2 digit
		1983-84 Grade Level	TH, SPAM	2 digit, 1-8 only
		1983-84 Free Lunch Status	SPAM	1 digit/1-3
		1983-84 Bilingual Fluency	SPAM	1 digit/1-3
		Chapter 1 Partic.	SPAM	1 digit/0,1
		Special Ed. Partic.	SPAM	1 digit/0,1
		Ethnic Group	SPAM	1 digit/1-5
		Sex	SPAM	1 digit/0,1
		Basal Reading Series	SPAM	1 digit
3.	PRI Performance	Fall, 1983 IOI test date	TRACER	4 digit/Julian
		Level of IOI admn.	TRACER	1 digit/1-5
		No. of instr. obj. tested on IOI	TRACER	2 digit/13-43
		No. of instr. obj. mastered on IOI	TRACER	2 digit/0-43
		No. of mastery tests administered during rest of year	TRACER	2 digit/0-99
		No. of objectives mastered via mastery tests during rest of year	TRACER	2 digits/0-99
		No. of objectives mastered via teacher rating during rest of year	TRACER	2 digits/0-99

Two changes:

- Repeat bracketted variables for each of five PRI levels
- Within each level, repeat all variables twice: that which occurred between Sept-April 15, and that occurring from April 15 to end of school year.

-RH

Table 3 -- Cont'd

Contents of a Student Data Record

Variable Cluster	Variable Name	Source	Field Length/ Possible Value
4. CAT Reading Test Data	Spring, 1984 Test No.	TH	3 digit/Testing Office Code
	Reading Vocabulary Raw Score	TH	3 digit
	Reading Vocabulary %ile	TH	2 digit/0-99
• Eliminate %ile score	Reading Vocabulary Scale Score	TH	3 digit
• Provide other three scores for math & language arts subtests & total tests -RJ	Reading Vocabulary NCE	TH	2 digit/0-99
	Reading Comprehension Raw Score	TH	3 digit
	Reading Comprehension %ile	TH	2 digit/0-99
	Reading Comprehension Scale Score	TH	3 digit
	Reading Comprehension NCE	TH	2 digit/0-99
	Total Reading Raw Score	TH	3 digit
	Total Reading %ile	TH	2 digit/0-99
	Total Reading Scale Score	TH	3 digit
	Total Reading NCE	TH	2 digit/0-99
	Spring 1983 Test No.	TH	3 digit/Testing Office Code
	Reading Vocabulary Raw Score	TH	3 digit
	Reading Vocabulary %ile	TH	2 digit/0-99
	Reading Vocabulary Scale Score	TH	3 digit
	Reading Vocabulary NCE	TH	2 digit/0-99
	Reading Comprehension Raw Score	TH	3 digit
	Reading Comprehension %ile	TH	2 digit/0-99
	Reading Comprehension Scale Score	TH	3 digit

Table 3 -- Cont'd

Contents of a Student Data Record

Variable Cluster	Variable Name	Source	Field Length/ Possible Value
	Reading Comprehension NCE	TH	2 digit/0-99
	Total Reading Raw Score	TH	2 digit/0-99
	Total Reading %ile	TH	2 digit/0-99
	Total Reading Scale Score	TH	2 digit/0-99
	Total Reading NCE	TH	2 digit/0-99

Activity 2: Test/Curriculum Match

The degree of fit between the content of the district reading curriculum and that of the PRI materials and California Achievement Test is an important issue in the evaluation of PRI use in SPS. These three sources of reading material have very different hierarchies of instructional objectives which represent their content. The PRI has as many as nine general skill clusters, each of which contains a range of category objectives which are further specified by instructional objectives in each level of its graded system. Since it is published by the same publisher, CTB/McGraw-Hill, the CAT has a similar content structure as the PRI, but certainly not identical. In general, the CAT provides a less specific outline of its content than the PRI--a common difference between achievement (CAT) and diagnostic (PRI) assessment tools. The match between these instruments at each grade or test level will be assessed in a rating scale of similarity developed by the authors.

The instructional objectives in the SPS K-6 reading curriculum are organized around eight general categories of objectives which are further specified by instructional objectives at each grade level. Most of these seem parallel to the PRI general skill clusters (e.g., Word Analysis), while others are at the level of PRI category objectives (e.g., Vocabulary). There are similar variations in structure between the CAT and SPS curriculum.

This activity will consist first of an analysis of the content hierarchies of each of these three systems by grade level. Their hierarchies will then be aligned on common levels of generality. Finally, the similarity of instructional objectives will be independently rated by two reading specialists

hired as consultants by NWREL. If there are different degrees of match at different grade levels, this will be noted in the interpretation of any grade-to-grade differences in other activities of the study.

Timelines/Data Needs

NWREL has all of the necessary content information on all levels of the CAT and PRI, and for grades K-6 in the SPS curriculum. Since the analysis in other study activities spans grades 1-8, the addition of SPS curriculum specification for grades 7 and 8 would allow for a rating of test/curriculum which covers the same student population as that of the other study activities. With this information, the authors could complete this study activity within two months of the final design approval.

Eliminate this activity - RG

Activity 3: PRI Implementation Survey

For purposes of the analysis of reading achievement in Activity 1, the level of use of the PRI materials will be represented simply by the number of objectives tested or mastered by each student after the fall IOI diagnostic test. While this may be an effective means of representing general levels of use, it does not provide any information as to teacher attitudes toward the materials, obstacles toward fuller implementation, etc. To obtain more specific information of this type, a survey of teachers and principals in the district will be conducted.

A mailed survey of this type has much to build upon. CTB/McGraw-Hill conducted a similar effort in its Learner Validation Studies of PRI use in four pilot districts across the U.S. A PRI review panel in SPS surveyed a sample of teachers and all elementary and middle school principals in the district at the end of the 1983-84 school year. Their findings and subsequent recommendations have provided helpful information to the current study design.

Before conducting another survey in SPS, the current authors propose to do some additional analysis of data already collected by the district's review panel. This secondary analysis will consist primarily of frequency breakdowns of responses by important characteristics of respondents (e.g., teaching level, basal reading series used, frequency of TRACER use, etc.). Some of these survey results may be contrasted with those obtained in other study activities. For example, teachers were asked to indicate the number of PRI objectives mastered by their students during the year. These results can be compared with those of the actual TRACER data analysis in Activity 1. If there are discrepancies, it may suggest that the use of the TRACER system is not necessarily snyonymous with the use of the PRI materials.

Following this reanalysis and the completion of TRACER/SPAM analysis in Activity 1, the authors will conduct another survey of a sample of SPS reading teachers and principals. Questions on this survey will include several which were on the 1983-84 survey. This will allow some comparison of PRI implementation across the two years. Questions will also be included which arise from the analysis in Activity 1. For example, if PRI use seems prevalent in some grades or with certain types of students more than others, the survey will attempt to elicit reasons as to why this is the case.

The sampling scheme used will be similar to that employed by the review panel. All elementary and middle school principals (N=80) and a random sample of three to five reading teachers per building (N=300) will be surveyed. Followup letters and telephone calls will be conducted to facilitate as high a response rate as possible. A cover letter signed by a designated district staff person would likely help in obtaining cooperation from building level staff throughout the district. Unlike the previous survey, however, the authors will attempt to collect identifying information from each respondent for research purposes in the study. In this way, survey findings may be linked with the data analysis of Activity 1 while still maintaining confidentiality of respondent identity.

Timelines/Data Needs

Secondary analysis of the 1983-84 survey data can begin as soon as the data are delivered to the authors. It will facilitate interpretation of the TRACER data analysis if the survey data can be made available before the data tape for Activity 1. The 1984-85 survey will be conducted near the end of the school year (May) as was the case in the previous year's survey.

Activity 4: SPS Staff Interviews

Interviewing a sample of teachers and principals is the final step along a continuum of specificity of implementation information in this study. Analysis of TRACER data will provide quantitative indicators of how frequently the PRI test instruments are used. Questionnaire data will provide more specific analysis of use of particular components of the PRI materials, the amount of time teachers spend with the system, and particular obstacles toward effective implementation. The interviews in this study are designed to fill in the details of these findings in a more narrative style which captures the actual flow of classroom activities and PRI use.

The selection of a sample of teachers and principals to interview has a different purpose than that of other study activities. It is not intended to be a representative sample of the entire district. Instead, it is designed to elaborate more fully on particularly effective practices in PRI use in SPS, as well as on those which make little or no use of the materials at all. Whether these extremes occur with representative frequency in the interview sample is not of concern to the authors. We hope to get enough representation of both extremes that a collection of in-depth interviews will yield informative details as to actual day-to-day experiences in the use of PRI materials.

Based on the analysis of TRACER data in Activity 1, the authors will have identified teachers at various extremes of PRI use. Using a roster of reading teachers at all elementary and middle schools in the district, the authors will select a sample of high, medium and low PRI use teachers (including those who don't use it at all and are not in the TRACER data base). Hopefully, the sample can be selected such that two or three teachers and the principal from the same building can be individually interviewed at the same school site. Care will be taken to ensure that an acceptable span of grade levels are

included. A minimum of six schools will ve visited during a week of interviewing, with at least 24 building level staff participating in the interviews. The authors will discuss the logistics of the interviewing with district and building level staff in advance to minimize disruption to ongoing school operations at any of the buildings.

Timelines/Data Needs

The authors will need a list of names, ID numbers and school mailing addresses of teachers who would be the designated PRI users, and the grade levels they teach, in all elementary and middle schools in the district. With this information, and following the initial TRACER analysis, interviews could be scheduled and conducted before Spring CAT testing.

SUMMARY

The specification of general timelines and data needs for all study activities is evidence of the authors intent and need to maintain ongoing communication with key SPS staff throughout the conduct of this evaluation. We propose a series of progress reports as deliverables in nearly each month of the contract. These brief reports will summarize results in the current study activity and will foster the ongoing communication between SPS and NWREL desired.The occurrence of these reports can be seen in the context of a list of "critical events" occurring during the course of the study. This list is presented as Table 4, with date of completion of each task noted, and indication of a progress report denoted by a series of asterisks (***). Nearly all of the evaluation activities planned by the authors are contingent upon receipt of the data tape from SPS which merges information from the SPAM, TRACER, and Test History files. This due date, as shown below, is February 1, 1985. If this timeline is not feasible for SPS, other due dates in the list of events also need to be adjusted accordingly.

Table 4

"Critical Events" in PRI Study

Task	Due Date	NWREL Deliverable
NWREL presents design to SPS	December 15, 1984	***
SPS reviews design	January 1, 1985	
Finalize design	January 4, 1985	
Activity 1: Data Analysis		
NWREL/SPS arrive at tape spec's	December 15, 1984	
SPS delivers tape to NWREL	February 1, 1985	
NWREL completes analysis of TRACER data	March 1, 1985	
NWREL Progress Report on analysis of PRI use	~~March 15, 1985~~ to May 15	***
SPS review report and contingencies for further analysis	March 22, 1985	
NWREL completes analysis of effects of PRI use on CAT scores for 1983-84 year	May 31, 1985	
Progress report of 1983-84 findings to SPS	June 15, 1985 → to Aug. 15	***
Finalize plans for longitudinal analysis	June 30, 1985	
Progress Report on longitudinal results to SPS	July 30, 1985	***

Table 4 -- Cont'd

"Critical Events" in PRI Study

Task	Due Date	NWREL Deliverable
Activity 2: Test/Curriculum Match		
Obtain grade 7-8 reading objectives from SPS	January 1, 1985	
NWREL develop comparison charts	January 15, 1985	
Complete ratings of PRI/CAT/SPS objectives	February 1, 1985	
Progress Report to SPS	~~February 15, 1985~~ to April 15	***
Activity 3: PRI Implementation Survey		
Obtain data from 1983-84 survey from SPS	January 1, 1985	
Prepare data for analysis	January 15, 1985	
Complete secondary analysis	February 15, 1985	
Progress Report of secondary analysis results to SPS	March 1, 1985	***
NWREL construct survey for 1984-85	March 31, 1985	
SPS review proposed survey	April 15, 1985	
NWREL conduct initial mail-out	May 1, 1985	
NWREL conduct first follow-up	May 15, 1985	
NWREL conduct second follow-up	May 30, 1985	
Complete analysis of survey data	June 15, 1985	
Progress Report to SPS on 1984-85 survey findings	June 30, 1985	***

Appendix B

Interview Guide and Schedule